A Short Decidability Proof for DPDA Language Equivalence via First-Order Grammars

Petr Jančar

Techn. Univ. Ostrava, Czech Republic,

http://www.cs.vsb.cz/jancar/, email:petr.jancar@vsb.cz

Abstract

The main aim of the paper is to give a short self-contained proof of the decidability of language equivalence for deterministic pushdown automata, which is the famous problem solved by G. Sénizergues, for which C. Stirling has derived a primitive recursive complexity upper bound. The proof here is given in the framework of first-order grammars, which seems to be particularly apt for the aim.

Keywords: pushdown automaton, deterministic context-free language, first-order grammar, language equivalence, trace equivalence, decidability

1 Introduction

The decidability question for language equivalence of two deterministic pushdown automata (dpda) is a famous problem in language theory. The question was explicitly stated in the 1960s [1] (when language inclusion was found undecidable); then a series of works solving various subcases followed, until the question was answered positively by Sénizergues in 1997 (a full version appeared in [2]). G. Sénizergues was awarded Gödel prize in 2002 for this significant achievement. Later Stirling [5], and also Sénizergues [3], provided simpler proofs than the original long technical proof. A modified version, which showed a primitive recursive complexity upper bound, appeared as a conference paper by Stirling in 2002 [6].

Nevertheless, even the above mentioned simplified proofs are rather technical, and they seem not well understood in the (theoretical) computer science community. One reason might be that the frameworks like that of strict deterministic grammars, which were chosen by Sénizergues and Stirling, are not ideal for presenting this topic to a broader audience.

From some (older) works by Courcelle, Harrison and others we know that the dpda-framework and strict-deterministic grammar framework are equivalent to the framework of first-order schemes, or first-order grammars. In this paper, a proof in the framework of first-order grammars is presented.

Author's remark. I have been reminded that J. R. Büchi in his book "Finite automata, their algebras and grammars: towards a theory of formal expressions" (1989) argues that using terms is the way proofs on context-free grammars should be done. I have not managed to verify myself, but this can be an indication that the framework of first-order terms might be "inherently more suitable" here.

In fact, the proof here shows the decidability of trace equivalence (a variant of language equivalence, coinciding with bisimulation equivalence on deterministic labelled transition systems) for deterministic first-order grammars; the states (configurations) are first-order terms which can change by performing actions according to the root-rewriting rules. To make the paper self-contained, a reduction from the dpda language equivalence problem to the above trace equivalence problem is given in an appendix.

In principle, the proof is lead in a similar manner as the proofs of Sénizergues and Stirling, being based on the same abstract ideas. Nevertheless, the framework of the first-order terms seems to allow to highlight the basic ideas in a more clear way and to provide a shorter proof in the form of a sequence of relatively simple observations. Though the proof is the "same" as the previous ones on an abstract level, and each particular idea used here might be embedded somewhere in the previous proofs, it is by no means a "mechanical translation" of a proof (or proofs) from one framework to another. Another appendix then shows that the framework chosen here also has a potential to concisely comprise and slightly strengthen the previous knowledge of the complexity of the problem, though the proofs in that part are not so detailed as in the decidability part.

Author's remarks. I hope that the presentation here should significantly extend the number of people in the community who will understand the problem which seems to belong to fundamental ones; this might also trigger new attempts regarding the research of complexity. Another remark concerns the previous version(s) of this arxiv-paper where I also claimed to provide a smooth generalization of the decidability proof to the case of bisimilarity for nondeterministic first-order grammars. Géraud Sénizergues was present at my talk at http://www.lsv.ens-cachan.fr/Events/Pavas/ (20 January, 2011) and he later put a counterexample on arxiv: http://arxiv.org/abs/1101.5046. (My mistake was, in fact, embarrassingly simple, mixing the absolute equivalence levels with the eq-levels relative to fixed strategies. At the moment, I do not speculate how this can be corrected.)

In the rest of this section, the main ideas are sketched. A GNF (Greibach Normal Form) grammar $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$, with finite sets $\mathcal{N}$ of nonterminals and $\mathcal{A}$ of terminals, has the rewriting rules $X \longrightarrow aY_1 \dots Y_n$ where $a \in \mathcal{A}$ and $X, Y_1, \dots, Y_n \in \mathcal{N}$. Such a rewrite rule can be written as $Xx \xrightarrow{a} Y_1 \dots Y_n x$, for a formal variable x, and read as follows: any sequence $X\alpha \in \mathcal{N}^*$ (a 'state', or a 'configuration') can perform action a while changing into $Y_1 \dots Y_n\alpha$; this includes the case when $n = 0$ and thus $Y_1 \dots Y_n = \varepsilon$, the empty word. The language $L(\alpha)$ is the set of words $w = a_1a_2 \dots a_m \in \mathcal{A}^*$ such that $\alpha \xrightarrow{a_1} .. \xrightarrow{a_2} .. \cdots .. \xrightarrow{a_m} \varepsilon$.

In a first-order grammar $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$, each nonterminal X has a finite arity (not only arity 1 like in the GNF grammar), and the (root rewriting) rules are $Xx_1 \dots x_m \xrightarrow{a} E(x_1, \dots, x_m)$ where $m = \mathit{arity}(X)$ and E is a finite term over $\mathcal{N}$ where all occurring variables are from the set $\{x_1, \dots, x_m\}$. When $G_1, \dots, G_m$ are terms then $XG_1 \dots G_m \xrightarrow{a} E(G_1, \dots, G_m)$ where $E(G_1, \dots, G_m)$ denotes the result of substitution $E(x_1, \dots, x_m)[G_1/x_1, \dots, G_m/x_m]$. Grammar $\mathcal{G}$ is *deterministic* if for each pair $X \in \mathcal{N}$, $a \in \mathcal{A}$ there is at most one rule of the type $Xx_1 \dots x_m \xrightarrow{a} \dots$. We note that states (configurations) are no longer strings (as in the case of GNF grammars) but *terms*, naturally represented as *trees*.

It is a routine to reduce dpda language equivalence to deterministic first-order grammar *trace equivalence*: two terms T, T' are equivalent, $T \sim T'$, iff the words (traces) $w \in \mathcal{A}^*$ enabled in T ($T \xrightarrow{w} T_1$ for some T_1) are the same as those enabled in T'. It is natural to define the equivalence-level $\textsc{EqLv}(T, T')$ as the maximal $k \in \mathbb{N}$ for which we have $T \sim_k T'$, which means that T, T' enable the same words upto length k; we put $\textsc{EqLv}(T, T') = \omega$ iff $T \sim T'$, i.e. iff $T \sim_k T'$ for all $k \in \mathbb{N}$.

Given a deterministic first-order grammar $\mathcal{G}$ and an initial pair of terms T_0, U_0, the first idea for deciding $T_0 \overset{?}{\sim} U_0$ is to use the (breadth-first) search for a shortest word w which is enabled in just one of T_0, U_0. We call such a *word* as *offending* (adopting the viewpoint of a defender of the claim $T_0 \sim U_0$).

To get a terminating algorithm in the case of $T_0 \sim U_0$, we can think of some *sound system* enabling to establish for certain words $u \in \mathcal{A}^*$ that they are not *offending prefixes*, i.e. prefixes of offending words for T_0, U_0; e.g., if $T_0 \xrightarrow{u} T$ and $U_0 \xrightarrow{u} T$ then u cannot be an offending prefix. We aim at *completeness*, i.e., look for some means which enable to recognize sufficiently many nonoffending prefixes, finally showing that ε is not offending, which means $T_0 \sim U_0$.

We use a simple observation that the eq-level of a pair (T, U) can drop by at most 1 when both sides perform the same action, and it really *drops by* 1 *in each step* when we follow an offending word. It is also easy to observe a *congruence property* of *subterm replacement*; in particular, given terms U_1, U_2 with $\textsc{EqLv}(U_1, U_2) = k$ and T_1, T_2 with $\textsc{EqLv}(T_1, T_2) \geq k + 1$ where U_1 has a subterm T_1, i.e. $U_1 = E(T_1)$, by replacing T_1 with T_2 we get for the arising $U_1' = E(T_2)$ that $\textsc{EqLv}(U_1', U_2) = \textsc{EqLv}(U_1, U_2) = k$; the eq-level has been unaffected, and moreover, even the offending words for (U_1', U_2) are the same as the offending words for (U_1, U_2).

The above simple observations allow to build a (sound and) complete system, when we add the notion of a *basis*, a finite set of pairs of 'equivalent heads' (tree-tops, tree-prefixes) $E(x_1, \ldots, x_n), F(x_1, \ldots, x_n)$, for which we have $E(T_1, \ldots, T_n) \sim F(T_1, \ldots, T_n)$ for every instance. To enable a smooth completeness proof, showing even the existence of a fixed sufficient basis $\mathcal{B}$ for each grammar $\mathcal{G}$ (not depending on the initial pair), it is helpful to start immediately in a more general setting of *regular terms*, which are finite or infinite terms with only finitely many subterms (where a subterm can possibly have an infinite number of occurrences); such terms have natural finite graph presentations.

The structure of the paper is clear from (sub)section titles. There are also two (above mentioned) appendices.

2 Basic definitions and simple facts

In this section we introduce the basic definitions and observe some simple facts on which the main proof is based.

By $\mathbb{N}$ we denote the set $\{0, 1, 2, \ldots\}$ of natural numbers; symbol ω is taken as the first

infinite ordinal number, which satisfies $n < \omega$ and $\omega - n = \omega$ for any $n \in \mathbb{N}$.
For a set A, by A^* we denote the set of finite sequences, i.e. words, of elements of A; the length of $w \in A^*$ is denoted $|w|$, and we use ε for the empty sequence, so $|\varepsilon| = 0$. By A^ω we denote the set of infinite sequences of elements of A (i.e., the mappings $\mathbb{N} \to A$).
Given $L \subseteq A^* \cup A^\omega$ and $u \in \mathcal{A}^*$, by $u \backslash L$ we mean the *left quotient* of L by u, i.e. the set $\{v \in \mathcal{A}^* \cup \mathcal{A}^\omega \mid uv \in L\}$. By $\text{PREF}(L) = \{u \in \mathcal{A}^* \mid uv \in L \text{ for some } v\}$ we denote the set of (finite) prefixes of the words in L.

(Deterministic) labelled transition systems and (stratified) trace equivalence

A *labelled transition system (LTS)* is a tuple $(\mathcal{S}, \mathcal{A}, (\xrightarrow{a})_{a\in\mathcal{A}})$ where $\mathcal{S}$ is the set of *states*, $\mathcal{A}$ the set of *actions* and $\xrightarrow{a} \subseteq \mathcal{S} \times \mathcal{S}$ is the set of *transitions labelled with a*.
We extend $\xrightarrow{a}$ to relations $\xrightarrow{w} \subseteq \mathcal{S} \times \mathcal{S}$ for all $w \in \mathcal{A}^*$ inductively: $s \xrightarrow{\varepsilon} s$; if $s \xrightarrow{a} s_1$ and $s_1 \xrightarrow{v} s_2$ then $s \xrightarrow{av} s_2$. We say that s' is *reachable from s by a word w* if $s \xrightarrow{w} s'$. A *state $s \in \mathcal{S}$ enables (a trace) $w \in \mathcal{A}^*$*, denoted $s \xrightarrow{w}$, if there is s' such that $s \xrightarrow{w} s'$. An infinite sequence $\alpha = a_1a_2a_3\ldots \in \mathcal{A}^\omega$ is enabled in s_0, written $s_0 \xrightarrow{\alpha}$, if there are $s_1, s_2, s_3, \ldots$ such that $s_i \xrightarrow{a_{i+1}} s_{i+1}$ for all $i \in \mathbb{N}$. The *trace-set* of a state $s \in \mathcal{S}$ is defined as

$$\text{TRAC}(s) = \{w \in \mathcal{A}^* \mid s \xrightarrow{w}\}.$$

For $k \in \mathbb{N}$, we define $\text{TRAC}^{\leq k}(s) = \{w \in \mathcal{A}^* \mid w \in \text{TRAC}(s) \text{ and } |w| \leq k\}$; we also put $\text{TRAC}^\omega(s) = \{\alpha \in \mathcal{A}^\omega \mid s \xrightarrow{\alpha}\}$. On the set $\mathcal{S}$, we define the *(trace) equivalence* $\sim$, and the *family of equivalences* $\sim_k$, $k \in \mathbb{N}$, as follows:

$$r \sim s \Leftrightarrow_{df} \text{TRAC}(r) = \text{TRAC}(s) \quad \text{and} \quad r \sim_k s \Leftrightarrow_{df} \text{TRAC}^{\leq k}(r) = \text{TRAC}^{\leq k}(s).$$

Observation 1 $\sim_0 = \mathcal{S} \times \mathcal{S}$. $\forall k \in \mathbb{N} : \sim_k \supseteq \sim_{k+1}$. $\bigcap_{k\in\mathbb{N}} \sim_k = \sim$.

The *equivalence level*, or the *eq-level*, of a pair of states is defined as follows:

$$\text{EqLv}(r, s) = k \text{ if } r \sim_k s \text{ and } r \not\sim_{k+1} s\,;\ \text{EqLv}(r, s) = \omega \text{ if } r \sim s.$$

The shortest words showing nonequivalence for a pair r, s (if $r \not\sim s$) are called *offending words*: $\text{OW}(r, s) = \{w \mid w \text{ is a shortest word in } (\text{TRAC}(r) \smallsetminus \text{TRAC}(s)) \cup (\text{TRAC}(s) \smallsetminus \text{TRAC}(r))\}$. The elements of $\text{PREF}(\text{OW}(r, s))$ are called *offending prefixes* for the pair (r, s). We now note some trivial facts, point 3 being of particular interest:

Observation 2 *(1.) $r \sim s$ iff $\text{OW}(r, s) = \emptyset$ iff $\varepsilon \notin \text{PREF}(\text{OW}(r, s))$.*
(2.) If $r \not\sim s$ then $\text{EqLv}(r, s) = |w| - 1$ for any $w \in \text{OW}(r, s)$.
(3.) If $\text{EqLv}(r, s) = k$ and $\text{EqLv}(r, q) \geq k + 1$ then
$\text{EqLv}(q, s) = k$ and $\text{OW}(r, s) = \text{OW}(q, s)$.

An *LTS* $(\mathcal{S}, \mathcal{A}, (\xrightarrow{a})_{a\in\mathcal{A}})$ is *deterministic* if each $\xrightarrow{a}$ is a partial function, i.e.: if $r \xrightarrow{a} s_1$ and $r \xrightarrow{a} s_2$ then $s_1 = s_2$. (Recall now the left quotient operation $u \backslash L$.)

Observation 3 *In any* deterministic *LTS, if* $r \xrightarrow{u} r'$ *then* $\textsc{trac}(r') = u\backslash\textsc{trac}(r)$*, and* $\textsc{trac}^{\omega}(r') = u\backslash\textsc{trac}^{\omega}(r)$*. Moreover,* $\textsc{trac}(r) = \textsc{trac}(s)$ *implies* $\textsc{trac}^{\omega}(r) = \textsc{trac}^{\omega}(s)$*.*

We use notation $(r,s) \xrightarrow{u} (r',s')$ as a shorthand meaning $r \xrightarrow{u} r'$ and $s \xrightarrow{u} s'$.

Proposition 4 *Assume a* deterministic *LTS, and suppose* $(r,s) \xrightarrow{u} (r',s')$*. Then:*
1. $\mathrm{E}\textsc{q}\mathrm{L}\textsc{v}(r,s) - |u| \leq \mathrm{E}\textsc{q}\mathrm{L}\textsc{v}(r',s')$ *(in particular,* $r \sim s$ *implies* $r' \sim s'$*);*
2. if $r \not\sim s$ *then* $\mathrm{E}\textsc{q}\mathrm{L}\textsc{v}(r,s) - |u| = \mathrm{E}\textsc{q}\mathrm{L}\textsc{v}(r',s')$ *iff* $u \in \textsc{pref}(\mathrm{OW}(r,s))$*;*
3. if $u \in \textsc{pref}(\mathrm{OW}(r,s))$ *then* $\mathrm{OW}(r',s') = u\backslash\mathrm{OW}(r,s)$*.*

Proof: This follows almost trivially from Observation 3. E.g., for Point 3. it is sufficient to note: if $u \in \textsc{pref}(\mathrm{OW}(r,s))$ then $uv \in \mathrm{OW}(r,s)$ iff $v \in \mathrm{OW}(r',s')$. □

Finite and infinite regular terms and their finite graph presentations

We now give (a variant of) standard definitions of first-order terms, including infinite terms; we fix a countable set $\mathcal{V} = \{x_1, x_2, x_3, \ldots\}$ of (first-order) *variables*.

Let us now assume a given *finite* set $\mathcal{N}$ of ranked symbols, called *nonterminals* (though we can also view them as function symbols). Each $X \in \mathcal{N}$ thus has $arity(X) \in \mathbb{N}$; we use X, Y to range over elements of $\mathcal{N}$.

A (general) *term* E over $\mathcal{N}$ (and $\mathcal{V}$) is defined as a partial mapping $E : \mathbb{N}^* \to \mathcal{N} \cup \mathcal{V}$ where the domain $\textsc{dom}(E) \subseteq \mathbb{N}^*$ is prefix-closed, i.e. $\textsc{dom}(E) = \textsc{pref}(\textsc{dom}(E))$, and nonempty ($\varepsilon \in \textsc{dom}(E)$); moreover, for $\gamma \in \textsc{dom}(E)$ we have $\gamma i \in \textsc{dom}(E)$ iff $1 \leq i \leq arity(E(\gamma))$ where the arity of variables $x_j \in \mathcal{V}$ is viewed as 0.

For each $\gamma \in \textsc{dom}(E)$, by $E_{[\gamma]}$ we denote *the subterm occurring* at γ in E where $E_{[\gamma]}(\delta) = E(\gamma\delta)$ for each $\delta \in \textsc{dom}(E_{[\gamma]}) = \gamma\backslash\textsc{dom}(E)$; this *occurrence of subterm* $E_{[\gamma]}$ has *depth* $|\gamma|$ in E.

For $X \in \mathcal{N}$ and terms $G_1, G_2, \ldots, G_m$, where $m = arity(X)$, by $XG_1G_2\ldots G_m$ we denote the term E for which $E(\varepsilon) = X$ and $E_{[i]} = G_i$ for each length-1 sequence i where $1 \leq i \leq m$; X is the *root nonterminal* of this term E. Each variable $x_j \in \mathcal{V}$ is also viewed as the term E for which $E(\varepsilon) = x_j$ (and thus $\textsc{dom}(E) = \{\varepsilon\}$).

A *term* E is *finite* (*infinite*) if $\textsc{dom}(E)$ is finite (infinite). The *depth-size* of a finite term E, denoted $\mathrm{D}\textsc{epth}(E)$, is the maximal $|\gamma|$ for $\gamma \in \textsc{dom}(E)$ (i.e., the maximal depth of a subterm-occurrence in E). A *term* is *regular* if the set of its subterms is finite (though the subterms can have infinitely many occurrences).

By $\mathrm{T}\textsc{erms}_{\mathcal{N}}$ we denote the set of all (finite and infinite) *regular* terms, since we will not consider nonregular terms anymore. Hence from now on, when saying "term" we mean "regular term". $\mathrm{GT}\textsc{erms}_{\mathcal{N}}$ denotes the set of all (regular) *ground terms*, i.e. the terms in which no variables x_i occur. We use symbols T, U, V, W (possibly with sub- and superscripts) for ranging over $\mathrm{GT}\textsc{erms}_{\mathcal{N}}$; symbols E, F, G, H are used more generally, they range over $\mathrm{T}\textsc{erms}_{\mathcal{N}}$.

Regular terms can be infinite but they have natural finite presentations, since they can be viewed as the unfoldings of finite graphs:

Definition 5 *A* graph presentation *of a regular term is a finite labelled (multi)graph g, where each node v has a label $\lambda(v) \in \mathcal{N} \cup \mathcal{V}$ and m outgoing edges labelled with $1, 2, \ldots, m$ where $m = arity(\lambda(v))$ (and where different edges can have the same target); moreover, one node is selected as the root. Graph g represents term $\mathcal{T}_g$ as follows:* $\textsc{dom}(\mathcal{T}_g)$ *consists of sequences of edge-labels of finite paths in g starting in the root; $\mathcal{T}_g(\gamma) = \lambda(v')$ where v' is the end-vertex of the path with the edge-label sequence γ.*

(Given $\mathcal{N}$,) we can naturally define a notion of the *size of a graph presentation* g, e.g., as the number of nodes of g, or, to be pedantic, as the length of a standard bit-string representation of g (thus also handling the descriptions of indexes of variables x_i). We define the *presentation size* of a term F, denoted $\textsc{PresSize}(F)$, as the size of the least graph presentation of F. By $\textsc{PresSize}(E, F)$ for a pair E, F we mean the sum $\textsc{PresSize}(E) + \textsc{PresSize}(F)$, say. On our level of reasoning, we do not need further technical details, since the facts like the following one are sufficient for us.

Observation 6
For any $s \in \mathbb{N}$, there are only finitely many pairs (E, F) with $\textsc{PresSize}(E, F) \leq s$.

We will also use some facts concerning an effective (algorithmic) work with finite presentations of regular terms. The next observation is an example of such a fact. Further we often leave such facts implicit.

Observation 7 *There is an algorithm which, given ($\mathcal{N}$ and) graph presentations g_1, g_2, decides if $\mathcal{T}(g_1) = \mathcal{T}(g_2)$.*

Substitutions, ground instances of a pair (E, F), a limit substitution

By a *substitution* we mean a mapping $\sigma : \mathcal{V} \rightarrow \textsc{Terms}_{\mathcal{N}}$. The term $E\sigma$ arises from E by replacing each occurrence x_i in E with $\sigma(x_i)$. (Since E and $\sigma(x_i)$ are regular, $E\sigma$ is regular.) By writing $\sigma = [G_{i_1}/x_{i_1}, \ldots, G_{i_n}/x_{i_n}]$ we mean that $\sigma(x_i) = G_i$ if $i \in \{i_1, \ldots, i_n\}$ and $\sigma(x_i) = x_i$ otherwise. Substitutions can be naturally composed; associativity allows to omit the parentheses: $E\sigma_1\sigma_2\sigma_3 = ((E\sigma_1)\sigma_2)\sigma_3 = E(\sigma_1(\sigma_2\sigma_3)) = (E\sigma_1)(\sigma_2\sigma_3)$, etc.
F is an *instance* of E if there is a substitution σ such that $E\sigma = F$.

As usual, we sometimes write $E(x_{i_1}, \ldots, x_{i_n})$ to denote the fact that all variables occurring in E are from the set $\{x_{i_1}, \ldots, x_{i_n}\}$. In fact, we only use the special case $E(x_1, \ldots, x_n)$; note that even a ground term E can be viewed as $E(x_1, \ldots, x_n)$, for any n. (We ignore the slight notational collision with the previous use $E(\gamma)$, since this should cause no problems.)

Convention. When writing $F(G_1, \ldots, G_n)$, we implicitly assume $F = F(x_1, \ldots, x_n)$, and we take $F(G_1, \ldots, G_n)$ as a shorthand for $F[G_1/x_1, \ldots, G_n/x_n]$. In particular we note that $F(G_1(H_1, \ldots, H_m), \ldots, G_n(H_1, \ldots, H_m)) = (F(G_1, \ldots, G_n))(H_1, \ldots, H_m)$.

A ground term U which is an instance of E is called a *ground instance* of E. We will in particular use the notion of a *ground instance of a pair* $(E(x_1, \ldots, x_n), F(x_1, \ldots, x_n))$: it is any pair $(E\sigma, F\sigma)$ where σ is a substitution $[U_1/x_1, \ldots, U_n/x_n]$ where U_i are ground terms. We usually write $(E(U_1, \ldots, U_n), F(U_1, \ldots, U_n))$ instead of $(E\sigma, F\sigma)$.

By writing $E\sigma^1$ we mean $E\sigma$; $E\sigma^{k+1}$ ($k \in \mathbb{N}$) means $E\sigma^k\sigma$. We need just a special case of substitutions σ for which $E\sigma^\omega$ is well-defined; we use the graph presentations for the definition (which also shows another aspect of the effective work with these finite presentations).

Definition 8 *Given (a regular term) H we define $H[H/x_i]^\omega$, denoted as H^{lim_i}, as follows: given a graph presentation g where $\mathcal{T}(g) = H$, then $H^{lim_i} = \mathcal{T}(g')$ where g' arises from g by redirecting each edge leading to a node labelled with x_i (in g) to the root (in g'). (Hence if $H = x_i$ or x_i does not occur in H then $H^{lim_i} = H$.)*

Head-tails presentations of terms, the d-prefix form of terms

If $F = E(G_1, \dots, G_n)$ then we say that the (regular) *head* E and the (regular) *tails* $G_1, \dots, G_n$ constitute a *head-tails presentation* of F. We can also note that the head E itself can be presented by a head-tails presentation $E = G(F_1, \dots, F_m)$, say, etc.
A particular head-tails presentation of a term is its d-prefix form; it suffices when we restrict ourselves to ground terms:

Definition 9 *For a ground term V and $d \in \mathbb{N}$, the d-*prefix form of V *arises as follows: we take all (ordered) occurrences of subterms of V with depth d (if any), say $T_1, \dots, T_n$, and replace them with variables $x_1, \dots, x_n$, respectively. We thus get a* finite *term $P_d^V = P_d^V(x_1, \dots, x_n)$, the d-*prefix of V. *The head P_d^V and the tails $T_1, \dots, T_n$ constitute the d-*prefix form *of $V = P_d^V(T_1, \dots, T_n)$. ($P_d^V = V$ when V is a finite term with* $\textsc{Depth}(V) < d$.*)*

Observation 10 *If m is the maximal arity of nonterminals in $\mathcal{N}$ then the number n of tails in the d-prefix form is bounded by m^d.*

We also note the next obvious fact.

Observation 11 *If $F = F(x_1) \neq x_1$ then F^{lim_1} is a ground term, and for any T we have $P_d^{F^{lim_1}} = P_d^{F\sigma^d(T)}$ where $\sigma = [F/x_1]$.*

First-order grammars as generators of LTSs

Definition 12 *A* first-order grammar *is a structure $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ where $\mathcal{N}$ is a finite set of ranked* nonterminals*, $\mathcal{A}$ is a finite set of* actions *(or terminals), and $\mathcal{R}$ a finite set of* (root rewriting) rules *of the form*

$$Xx_1x_2\dots x_m \xrightarrow{a} E(x_1, x_2, \dots, x_m) \tag{1}$$

where $X \in \mathcal{N}$, $arity(X) = m$, $x_1, x_2, \dots, x_m \in \mathcal{V}$, $a \in \mathcal{A}$, and E is a finite *term over $\mathcal{N}$ (and $\mathcal{V}$) in which all occurring variables are from the set $\{x_1, x_2, \dots, x_m\}$. ($E = x_i$ is a particular example.)* Grammar *$\mathcal{G}$ is* deterministic*, a* det-first-order grammar*, if for each pair $X \in \mathcal{N}$, $a \in \mathcal{A}$ there is at most one rule of the form (1).*

Remark. Context-free grammars in Greibach normal form can be seen as a special case, where each nonterminal has arity 1. Classical rules like $A \to aBC$, $B \to b$ can be presented as $Ax_1 \xrightarrow{a} BCx_1$, $Bx_1 \xrightarrow{b} x_1$.

We view $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ as a generator of the $LTS_{\mathcal{G}} = (\textsc{Terms}_{\mathcal{N}}, \mathcal{A}, (\xrightarrow{a})_{a\in\mathcal{A}})$

where each (root) rewriting rule $Xx_1x_2\dots x_m \xrightarrow{a} E(x_1, x_2, \dots, x_m)$ is a "schema" (a "template") of a set of transitions: for every substitution $\sigma = [G_1/x_1, \dots, G_m/x_m]$ (including σ with $\sigma(x_i) = x_i$ for all i) we have $(Xx_1\dots x_m)\sigma \xrightarrow{a} E(x_1, \dots, x_m)\sigma$, i.e.

$$XG_1G_2\dots G_m \xrightarrow{a} E(G_1, G_2, \dots, G_m).$$

Observation 13 *When $\mathcal{G}$ is deterministic then $LTS_{\mathcal{G}}$ is deterministic.*

Though the main result applies to deterministic grammars, we will also note some properties holding in the general (nondeterministic) case.

The notions and notation like $F \xrightarrow{w}$ (w is enabled by F), $F \xrightarrow{w} F'$ (for words $w \in \mathcal{A}^*$), $F \xrightarrow{\alpha}$ (for $\alpha \in \mathcal{A}^\omega$), $\textsc{trac}(F)$, $\textsc{trac}^{\leq k}(F)$, $\textsc{trac}^{\omega}(F)$, trace equivalence $E \sim F$, etc., are inherited from $LTS_{\mathcal{G}}$. (Note that the term x_i enables no actions; nevertheless, it is technically convenient to have also nonground terms as states in $LTS_{\mathcal{G}}$.)
We will be interested in comparing ground terms, so we will use $T \sim_k U$, $T \sim U$, $\textsc{EqLv}(T,U)$, $\mathrm{OW}(T,U)$, $\textsc{pref}(\mathrm{OW}(T,U))$ mainly for $T, U \in \mathrm{GTerms}_{\mathcal{N}}$. We note (now explicitly) another fact about the effective work with graph presentations:

Observation 14 *Given ($\mathcal{G}$ and) a graph presentation g (of term $\mathcal{T}(g)$), we can effectively find all rewriting rules in $\mathcal{R}$ (of type (1)) which can be applied to $\mathcal{T}(g)$ and for each such rule yielding $\mathcal{T}(g) \xrightarrow{a} F$ we can effectively construct some g' such that $\mathcal{T}(g') = F$.*

Root-locality of action performing, words exposing subterm occurrences

Assuming a given first-order grammar $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$, we observe some consequences of the fact that the root nonterminal of $XG_1\dots G_m$ determines if $a \in \mathcal{A}$ is enabled, and when a transition is performed then the subterms $G_1, \dots, G_m$ play no role other than being copied (or "lost") appropriately.

Observation 15 *For $a \in \mathcal{A}$ and $u \in \mathcal{A}^*$, we have $au \in \textsc{trac}(XG_1\dots G_m)$ iff there is a rule $Xx_1\dots x_m \xrightarrow{a} E(x_1, \dots, x_m)$ in $\mathcal{R}$ such that $u \in \textsc{trac}(E)$ or $u = u_1u_2$ where $E \xrightarrow{u_1} x_i$ and $u_2 \in \textsc{trac}(G_i)$ (for some $i, 1 \leq i \leq m$).*

We say that $w \in \mathcal{A}^*$ *exposes the i-th successor* of $X \in \mathcal{N}$ if $Xx_1\dots x_m \xrightarrow{w} x_i$.

Observation 16 *Viewing a (regular) term E as a partial mapping $E : \mathbb{N}^* \to \mathcal{N} \cup \mathcal{V}$, we note that there is u such that $E \xrightarrow{u} x_i$ iff there is $\gamma \in \textsc{dom}(E)$ where $E(\gamma) = x_i$ and for each prefix δj of γ there is some w exposing the j-th successor of $E(\delta)$.*

Observation 17 $\textsc{trac}(E(G_1,\dots,G_n)) =$
$\textsc{trac}(E(x_1,\dots,x_n)) \cup \bigcup_{1\leq i\leq n}\{uv \mid E \xrightarrow{u} x_i \text{ and } v \in \textsc{trac}(G_i)\}$.

We now look at some simple algorithmic consequences of the above observations.

Proposition 18 *There is an algorithm which, given $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$, computes (and fixes) a word $w(X,i)$ for each pair (X,i), where $X \in \mathcal{N}$, $1 \leq i \leq arity(X)$, so that $w(X,i)$ is a shortest word exposing the i-th successor of X if any such word exists and $w(X,i) = \varepsilon$ otherwise.*

Proof: Recalling Observations 15 and 16, a brute-force systematic search of all $w(X,i)$ is sufficient: we finish when finding that the remaining pairs (X,i), i.e. those for which $w(X,i)$ have not been computed, are mutually dependent, i.e., the existence of an exposing $w(X,i)$ for any of the remaining pairs depends on the existence of exposing words for some remaining pairs. □

For later use we note that we can compute a bound bigger than any $|w(X,i)|$, say

$$M_0 = 1 + \max\{\, |w(X,i)| \mid X \in \mathcal{N}, 1 \leq i \leq arity(X)\,\}\,. \tag{2}$$

We also note the bounded increase of the depth-size of finite terms, given by the fact that the right-hand sides of the (finitely many) rules (1) in $\mathcal{R}$ are finite terms.

Observation 19 *(For $\mathcal{G}$,) there is a (linear) nondecreasing function* $\textsc{B-Inc} : \mathbb{N} \to \mathbb{N}$ *such that: if $F \xrightarrow{u} F'$ for a finite term F then* $\textsc{Depth}(F') \leq \textsc{Depth}(F) + \textsc{B-Inc}(|u|)$.

Generally we cannot provide a similar *lower* bound for $\textsc{Depth}(F')$, since some x_i might not occur in E in (1). But we can recall the d-prefix form from Definition 9 (defined for ground terms) and note the following obvious fact.

Observation 20 *If $V = P^V_d(T_1,\dots,T_n)$ and $|u| \leq d$ then $V \xrightarrow{u} V'$ iff there is some (finite) E such that $P^V_d(x_1,\dots,x_n) \xrightarrow{u} E(x_1,\dots,x_n)$ and $V' = E(T_1,\dots,T_n)$ (where* $\textsc{Depth}(E) \leq d + \textsc{B-Inc}(|u|)$*). Hence if $P^U_d = P^V_d$ then $U \sim_d V$.*

Congruence property of $\sim_k$ and $\sim$

Proposition 21 *If $T \sim_k T'$ then $E(T) \sim_k E(T')$. ($T \sim T'$ implies $E(T) \sim E(T')$.)*

Proof: The claim follows from Observation 17: $\textsc{trac}(E(T))$ consists of traces $w \in \textsc{trac}(E(x_1))$ and of traces of the form $w = uv$ where $E(x_1) \xrightarrow{u} x_1$ and $v \in \textsc{trac}(T)$. □

Proposition 22 *If $T \sim_k F(T)$ where $F \neq x_1$ then $T \sim_k F^{lim_1}$.*
Hence $T \sim F(T)$ implies $T \sim F^{lim_1}$.

Proof: If $T \sim_k F(T)$ then by repeated use of Proposition 21 we get $T \sim_k F\sigma^k(T)$ where $\sigma = [F/x_1]$. By Observation 11 we get $P_k^{F\sigma^k(T)} = P_k^{F^{lim_1}}$ and thus $F\sigma^k(T) \sim_k F^{lim_1}$ (by Observation 20), which implies $T \sim_k F^{lim_1}$. □

Corollary 23 *If $T_i \sim_k H(T_1, \ldots, T_n)$ where $H \neq x_i$ then $T_i \sim_k H^{lim_i}(T_1, \ldots, T_n)$ (with no occurrence of x_i in H^{lim_i}). In particular, if $T_n \sim_k H(T_1, \ldots, T_n)$ where $H \neq x_n$ then $T_n \sim_k H^{lim_n}(T_1, \ldots, T_{n-1})$.*

A normal form for first-order grammars

Observation 24
If there is no u such that $E(x_1) \xrightarrow{u} x_1$ then $E(T) \sim E(T')$ for any T, T'.

Definition 25 *$\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ is in* normal form *if for each $X \in \mathcal{N}$ and $i, 1 \leq i \leq arity(X)$ there is (a shortest) $w(X, i)$ exposing the i-th successor of X.*

Proposition 26 *Any $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ can be effectively transformed to $\mathcal{G}' = (\mathcal{N}', \mathcal{A}, \mathcal{R}')$ in normal form, yielding also an effective mapping* $\textsc{trans} : \textsc{Terms}_{\mathcal{N}} \rightarrow \textsc{Terms}_{\mathcal{N}'}$ *such that $T \sim \textsc{trans}(T)$ (in the disjoint union of $LTS_{\mathcal{G}}$ and $LTS_{\mathcal{G}'}$).*

Proof: For each $Y \in \mathcal{N}$, by $ES_Y = (i_1^Y, i_2^Y, \ldots, i_{k_Y}^Y)$ (Exposable Successors of Y) we denote the subsequence of $(1, 2, \ldots, arity(Y))$ where $i \in ES_Y$ iff there is w exposing the i-th successor of Y; by Y' we denote a fresh nonterminal with $arity(Y') = k_Y$. We put $\textsc{trans}(x_j) = x_j$ and $\textsc{trans}(YG_1 \ldots G_m) = Y' \, \textsc{trans}(G_{i_1^Y}) \ldots \textsc{trans}(G_{i_{k_Y}^Y})$. Each rule $Xx_1 \ldots x_m \xrightarrow{a} E(x_1, \ldots, x_m)$ in $\mathcal{R}$ is transformed to the rule $X'x_1 \ldots x_{k_X} \xrightarrow{a} \textsc{trans}(E)\sigma$ where $\sigma = [x_1/x_{i_1^X}, \ldots, x_{k_X}/x_{i_{k_X}^X}]$; thus $\mathcal{R}'$ arises. This guarantees that $Yx_1 \ldots x_m \xrightarrow{u} x_{i_j^Y}$ iff $Y'x_1 \ldots x_{k_Y} \xrightarrow{u} x_j$. The claim now follows from Proposition 18 and Observation 24. □

Exposing equations for pairs (E, F) in deterministic $LTS_{\mathcal{G}}$

We now note an important notion and make some observations.

Definition 27 *We say that $u \in \mathcal{A}^*$* exposes an equation for the pair *$(E(x_1, \ldots, x_n), F(x_1, \ldots, x_n))$ if $E(x_1, \ldots, x_n) \xrightarrow{u} x_i$, $F(x_1, \ldots, x_n) \xrightarrow{u} H(x_1, \ldots, x_n)$, or vice versa, where $H(x_1, \ldots, x_n) \neq x_i$ (but might be $H = x_j$ for $i \neq j$).
Formally we write this exposed equation as $x_i \doteq H(x_1, \ldots, x_n)$.*

Proposition 28 *For deterministic $\mathcal{G}$, if $E(T_1', \ldots T_n') \sim_k F(T_1', \ldots, T_n')$ and $E(T_1, \ldots T_n) \not\sim_k F(T_1, \ldots, T_n)$ then an offending prefix for $(E(T_1, \ldots, T_n), F(T_1, \ldots, T_n))$ exposes an equation for E, F.*

Proof: There is surely no $u \in \text{PREF}(\text{OW}(E(T_1,\dots,T_n),F(T_1,\dots,T_n)))$ such that $(E,F) \xrightarrow{u} (x_i,x_i)$. If there is $wa \in \text{OW}(E(T_1,\dots,T_n),F(T_1,\dots,T_n))$ (where $|w| < k$) such that $(E,F) \xrightarrow{w} (E',F')$ where none of E',F' is a variable and $E' \not\sim_1 F'$ then wa witnesses that $E(T'_1,\dots,T'_n) \not\sim_k F(T'_1,\dots,T'_n)$ – a contradiction. □

We recall that Observation 3 and Proposition 4 apply to $LTS_{\mathcal{G}}$ when $\mathcal{G}$ is deterministic, and observe the following:

Observation 29 *For deterministic $\mathcal{G}$, if u exposes an equation $x_i \doteq H(x_1,\dots,x_n)$ for E,F then* $\text{EqLv}(E(T_1,\dots,T_n),F(T_1,\dots,T_n)) - |u| \leq \text{EqLv}(T_i,H(T_1,\dots,T_n))$ *(for any $T_1,\dots,T_n$). Thus if $E(T_1,\dots,T_n) \sim F(T_1,\dots,T_n)$ then $T_i \sim H^{lim_i}(T_1,\dots,T_n))$; otherwise (at least) $T_i \sim_k H^{lim_i}(T_1,\dots,T_n))$ for $k =$* $\text{EqLv}(E(T_1,\dots,T_n),F(T_1,\dots,T_n)) - |u|$.

3 A "word-labelling predicate" ⊨ and its soundness

Definition 30 *(Given $\mathcal{G}$,) a* pair *$(E(x_1,\dots,x_n),F(x_1,\dots,x_n))$ is* sound *if we have $E\sigma \sim F\sigma$ for all ground instances $(E\sigma,F\sigma)$ (i.e., $(E(T_1,\dots,T_n),F(T_1,\dots,T_n))$) of the pair (E,F). A* set *$\mathcal{B}$ of* pairs *(E,F) is* sound *if each element is sound.*
By $\text{GInst}(\mathcal{B})$ *we mean the set of all ground instances of pairs in $\mathcal{B}$.*

Let us now assume a given *deterministic* first-order grammar $\mathcal{G} = (\mathcal{N},\mathcal{A},\mathcal{R})$ in normal form, and a *finite* set $\mathcal{B}$, called a *basis*, of pairs (of regular terms) (E,F), supposedly sound; we always assume $\mathcal{B}$ contains the pair (x_1,x_1) which is obviously sound.

The derivation (or deduction) system in Figure 1, assuming $\mathcal{G}$ and $\mathcal{B}$, provides an inductive definition of the predicate

$$\models_{(T_0,U_0)} \subseteq \mathcal{A}^* \times ((\text{GTerms}_{\mathcal{N}} \times \text{GTerms}_{\mathcal{N}}) \cup \{\text{NOP},\text{FAIL}\}).$$

So it is in fact a family of predicates, parametrized with the initial pair (T_0,U_0) of regular *ground* terms. We write just $\models$ instead of $\models_{(T_0,U_0)}$ when (T_0,U_0) is clear from context.

We can see that Axiom, Basic transition rule (1.) and Rejection rule (6.) guarantee that if $T_0 \not\sim U_0$ and ua is an offending word for (T_0,U_0), which implies $(T_0,U_0) \xrightarrow{u} (T,U)$ where $T \not\sim_1 U$, then $u \models (T,U)$, which can be read as "u can be labelled with the pair (T,U)", and thus $\varepsilon \models \text{FAIL}$; as expected, $\varepsilon \models \text{FAIL}$ is intended to mean $T_0 \not\sim U_0$.

In the case $T_0 \sim U_0$ we are guaranteed that if $(T_0,U_0) \xrightarrow{u} (T,U)$ then $u \models (T,U)$ but it is not clear how to use this to conclude that $T_0 \sim U_0$. To this aim we introduce another "label": $u \models \text{NOP}$ (read "u can be (also) labelled with NOP") is intended to imply that u is Not an Offending Prefix for (T_0,U_0) if $\mathcal{B}$ is sound. Thus $\varepsilon \models \text{NOP}$ is intended to imply $T_0 \sim U_0$ if $\mathcal{B}$ is sound. Having this in mind, Basis rule (4.) is clear; we will later realize the reason for the condition $u \neq \varepsilon$. Bottom-up progression rule (5.) is clear as well; we note in particular that it enables to derive $u \models \text{NOP}$ when $u \models (T,U)$ and T,U do not enable any action.

The most interesting is the Limit subterm replacement rule (2.), with its particular case of Subterm replacement rule. It allows to label u also with other pairs than those

- (Axiom) $\varepsilon \models (T_0, U_0)$
- (Primary derivation (deduction) rules, determining when $u \models (T, U)$)
  1. (Basic transition)
     If $u \models (T, U)$, $T \sim_1 U$, $(T, U) \xrightarrow{a} (T', U')$ then $ua \models (T', U')$.
  2. (Limit subterm replacement)
     If $u \models (E(T), U)$, $|v| < |u|$, $v \models (T, F(T))$, $F \neq x_1$ then $u \models (E(F^{lim_1}), U)$.
     (A particular case is Subterm replacement: if $v \models (T, T')$ then $u \models (E(T'), U)$.)
  3. (Symmetry) If $u \models (T, U)$ then $u \models (U, T)$.
- (Secondary derivation rules, determining when $u \models$ NOP and/or $u \models$ FAIL)
  4. (Basis) If $u \neq \varepsilon$, $u \models (T, U)$ and $(T, U) \in \mathrm{GInst}(\mathcal{B})$ then $u \models$ NOP.
  5. (Bottom-up progression) If $u \models (T, U)$, $T \sim_1 U$, and $ua \models$ NOP for all a enabled by T (and U) then $u \models$ NOP.
  6. (Rejection) If $u \models (T, U)$ where $T \not\sim_1 U$ then $\varepsilon \models$ FAIL.

Figure 1: Inductive definition of $\models_{(T_0,U_0)}$ (for given $\mathcal{G}, \mathcal{B}$)

derived just by Basic transition; so one u can get many pairs of terms as "labels". This is meant to help to create instances of the basis and label the respective words with NOP. The condition $|v| < |u|$ is important for soundness of the predicate $\models$ (wrt its intended meaning). The symmetry rule (3.) could be dropped if we included all symmetric cases in the Limit subterm replacement rule.

As usual, we write $u \models (T, U)$, or $u \not\models (T, U)$, if the predicate is true (i.e., derivable) for the triple u, T, U, or not true (not derivable), respectively; similarly for NOP and FAIL.

We now recall Observation 3 and Proposition 4, and show the following generalization in our case of det-first-order grammars. This is the crucial point for showing soundness (Proposition 33). (In fact, Point 3. is used later in the completeness proof.)

Proposition 31 *Given (a det-first-order grammar $\mathcal{G}$ and) an initial pair (T_0, U_0):*
1. If $u \models (T, U)$ then $\mathrm{EqLv}(T_0, U_0) - |u| \leq \mathrm{EqLv}(T, U)$. ($T_0 \sim U_0$ implies $T \sim U$.)
2. If $T_0 \not\sim U_0$ and $u \models (T, U)$ where $u \in \mathrm{PREF}(\mathrm{OW}(T_0, U_0))$ then
$\mathrm{EqLv}(T_0, U_0) - |u| = \mathrm{EqLv}(T, U)$ and $\mathrm{OW}(T, U) = u \backslash \mathrm{OW}(T_0, U_0)$.
3. If $T_0 \sim U_0$ and $u \models (T, U)$ then $\mathrm{TRAC}(T) = \mathrm{TRAC}(U) = u \backslash \mathrm{TRAC}(T_0) = u \backslash \mathrm{TRAC}(U_0)$.

Proof: We proceed by induction on the length of derivations. The axiom $\varepsilon \models (T_0, U_0)$ trivially satisfies the conditions. Suppose that the conditions are satisfied for all $u \models (T, U)$ derived by derivations upto length m, and consider a derivation deriving $u' \models (T', U')$ by $m + 1$ applications of the derivation rules.

If the last rule was 1. (Basic transition), using $u \models (T,U)$ and $(T,U) \xrightarrow{a} (T',U')$, then $ua \models (T',U')$ satisfies the conditions by (the induction hypothesis and) Propositions 3, 4: We have $\text{EqLv}(T,U) - 1 \leq \text{EqLv}(T',U')$. If $ua \in \text{PREF}(\text{OW}(T_0,U_0))$ then $u \in \text{PREF}(\text{OW}(T_0,U_0))$, hence $\text{OW}(T,U) = u \backslash \text{OW}(T_0,U_0)$ and thus $a \in \text{PREF}(\text{OW}(T,U))$; this implies $\text{OW}(T',U') = a \backslash \text{OW}(T,U)$ and thus $\text{OW}(T',U') = (ua) \backslash \text{OW}(T_0,U_0)$.

If the last rule was Limit subterm replacement (2.), so from $u \models (E(T),U)$, $|v| < |u|$, $v \models (T,F(T))$, $F \neq x_1$ we have derived $u \models (E(F^{lim_1}),U)$, then the conditions 1. and 3. (for $u \models (E(F^{lim_1}),U)$) follow from (the induction hypothesis and) Propositions 22 and 21; the condition 2. follows from Point 3. of Observation 2.

Rule 3. (Symmetry) obviously preserves the conditions. □

Corollary 32 *$T_0 \not\sim U_0$ iff $\varepsilon \models \text{FAIL}$.*

Proposition 33 *(Soundness)*
If $\varepsilon \models_{(T,U)} \text{NOP}$ for all $(T,U) \in \{(T_0,U_0)\} \cup \text{GInst}(\mathcal{B})$ then $\mathcal{B}$ is sound and $T_0 \sim U_0$.

Proof: By contradiction.
Suppose the assumption holds but there is some $(T_0',U_0') \in \{(T_0,U_0)\} \cup \text{GInst}(\mathcal{B})$ with the least finite eq-level. Take the longest (offending) prefix u of some $w \in \text{OW}(T_0',U_0')$ such that $u \models_{(T_0',U_0')} \text{NOP}$. The rule deriving $u \models_{(T_0',U_0')} \text{NOP}$ could not be the Basis rule since this supposes $u \neq \varepsilon$ and $u \models_{(T_0',U_0')} (T,U)$ where $(T,U) \in \text{GInst}\mathcal{B}$ but Point 2. of Proposition 31 implies that $\text{EqLv}(T,U)$ is smaller than the eq-level of any pair in $\text{GInst}(\mathcal{B})$. The same Point 2. also implies that the deriving rule could not be Bottom-up progression since this presupposes $ua \models_{(T_0',U_0')} \text{NOP}$ for a longer prefix ua of the above $w \in \text{OW}(T_0',U_0')$. □

So we have a sound system, on condition $\mathcal{B}$ is sound. But we note that an algorithm surely cannot process infinitely many pairs $(T,U) \in \text{GInst}(\mathcal{B})$ (to show $\varepsilon \models_{(T,U)} \text{NOP}$ for each of them). Fortunately, it suffices to consider a "critical instance" for each pair (E,F) in $\mathcal{B}$ which has the least eq-level among the ground instances of (E,F):

Definition 34 *The* critical instance *of a pair $(E(x_1,\ldots,x_n), F(x_1,\ldots,x_n))$ is the pair $(E\sigma, F\sigma)$ where $\sigma = [L_1/x_1,\ldots,L_n/x_n]$ for fresh nullary nonterminals L_i (extending $\mathcal{G}$) such that $L_i \not\sim_1 V$ for any $V \neq L_i$; e.g., L_i gets its own special action ℓ_i and a rule $L_i \xrightarrow{\ell_i} L_i$. By $\text{CritInst}(\mathcal{B})$ we mean the set of the critical instances of the pairs in $\mathcal{B}$.*

Proposition 35 *For any $E(x_1,\ldots,x_n)$, $F(x_1,\ldots,x_n)$, and any $T_1,\ldots,T_n$,*
$\text{EqLv}(E(L_1,\ldots,L_n),F(L_1,\ldots,L_n)) \leq \text{EqLv}(E(T_1,\ldots,T_n),F(T_1,\ldots,T_n))$
if $L_1,\ldots,L_n$ do not occur in E,F.

Proof: Proposition 28 shows that if we had $k_1 = \text{EqLv}(E(L_1,\ldots,L_n),F(L_1,\ldots,L_n)) > k_2 = \text{EqLv}(E(T_1,\ldots,T_n),F(T_1,\ldots,T_n))$ then there were an offending prefix u for $(E(T_1,\ldots,T_n),F(T_1,\ldots,T_n))$ exposing an equation $x_i \doteq H(x_1,\ldots,x_n)$ for E,F, which also means $k_2 \geq |u|$; but then $(E(L_1, \ldots, L_n), F(L_1, \ldots, L_n)) \xrightarrow{u} (L_i, H(L_1, \ldots, L_n))$ where $L_i \not\sim_1 H(L_1, \ldots, L_n)$, and thus $k_1 \leq |u|$ (a contradiction). □

The next lemma summarizes some important ingredients for the decidability of trace equivalence for det-first-order grammars, showing that it is now sufficient to prove the completeness of $\models$.

Lemma 36 *For a det-first-order grammar $\mathcal{G}$ and a finite set $\mathcal{B}$ of pairs of terms, if*

$$\varepsilon \models_{(T,U)} \mathrm{NOP} \textit{ for all } (T,U) \in \{(T_0,U_0)\} \cup \textsc{CritInst}(\mathcal{B}) \tag{3}$$

then $\mathcal{B}$ is sound and $T_0 \sim U_0$. Moreover, the condition (3) *is semidecidable.*

4 Completeness of the predicate $\models$

Definition 37 *Given a det-first-order grammar $\mathcal{G}$, a finite set $\mathcal{B}$ of pairs of regular terms is a* sufficient basis *if $\mathcal{B}$ is sound and we have: if T_0, U_0 are regular ground terms where $T_0 \sim U_0$ then $\varepsilon \models_{(T_0,U_0)}$ NOP (wrt $\mathcal{G}, \mathcal{B}$).*

We now aim to prove that any det-first-order grammar (in normal form) has a sufficient basis. We use implicitly the fact that if $T_0 \sim U_0$ then $u \models_{(T_0,U_0)} (T,U)$ implies $T \sim U$, $\textsc{trac}(T) = \textsc{trac}(U) = u\backslash\textsc{trac}(T_0) = u\backslash\textsc{trac}(U_0)$, and $\textsc{trac}^\omega(T) = \textsc{trac}^\omega(U) = u\backslash\textsc{trac}^\omega(T_0) = u\backslash\textsc{trac}^\omega(U_0)$ (recall Propositions 31 and 3). We start with a simple observation:

Proposition 38 *Given a det-first-order grammar $\mathcal{G}$ and a sound basis $\mathcal{B}$, if for every triple T_0, U_0, α where $T_0 \sim U_0$ and $\alpha \in \textsc{trac}^\omega(T_0)$ ($= \textsc{trac}^\omega(U_0)$) there is a prefix u of α such that $u \models_{(T_0,U_0)}$ NOP then $\mathcal{B}$ is a sufficient basis.*

Proof: Suppose $T_0 \sim U_0$. For every maximal $w \in \textsc{trac}(T_0)$ (for which there is no $wa \in \textsc{trac}(T_0)$) we have $w \models_{(T_0,U_0)}$ NOP by using Basic transition and Bottom-up progression. Thus the assumption that each $\alpha \in \textsc{trac}^\omega(T_0)$ has a prefix u for which $u \models_{(T_0,U_0)}$ NOP implies that $\varepsilon \models_{(T_0,U_0)}$ NOP, by repeated use of Bottom-up progression. □

We now show a sufficient condition for the existence of a sufficient basis (Proposition 42), first introducing some auxiliary notions to this aim.

Definition 39 *We assume a det-first-order grammar $\mathcal{G}$. Given (T_0, U_0) where $T_0 \sim U_0$, an infinite trace $\alpha \in \textsc{trac}^\omega(T_0)$ is* s-bounded *(for $s \in \mathbb{N}$) if it has a nonempty prefix u such that $u \models_{(T_0,U_0)} (E(T_1, \ldots, T_n), F(T_1, \ldots, T_n))$ where the pair $(E(x_1, \ldots, x_n), F(x_1, \ldots, x_n))$ is sound and $\textsc{PresSize}(E, F) \leq s$.*
Grammar $\mathcal{G}$ *is* s-bounded *if for each $T_0 \sim U_0$ all $\alpha \in \textsc{trac}^\omega(T_0)$ are s-bounded.*
Grammar $\mathcal{G}$ has a stair-base of width $n \in \mathbb{N}$ *if there is a function $g : \mathbb{N} \to \mathbb{N}$ such that: for every $T_0 \sim U_0$ and every $\alpha \in \textsc{trac}^\omega(T_0)$ either α is $g(0)$-bounded or there are some (unspecified) terms $T_1, \ldots, T_n$ and infinitely many nonempty, increasing prefixes $w_0, w_1, w_2, \ldots$ of α such that $w_j \models_{(T_0,U_0)} (E_j(T_1, \ldots, T_n), F_j(T_1, \ldots, T_n))$ for some (regular) E_j, F_j with $\textsc{PresSize}(E_j, F_j) < g(j)$, for all $j = 0, 1, 2, \ldots$.*

For illustration and later use, we first note a particular example of s-bounded traces where $s = \textsc{PresSize}(x_1, x_1)$; then we show that the stair-base property is indeed sufficient.

Definition 40 *Given $\mathcal{G}$ and an initial pair (T_0, U_0), $w \in \mathcal{A}^* \cup \mathcal{A}^\omega$* has a repeat *if there are two different prefixes u_1, u_2, $|u_1| < |u_2|$, of w and some T, U such that $u_1 \models (T, U)$, $u_2 \models (T, U)$. (By Subterm replacement we then derive $u_2 \models (U, U)$, where $(U, U) = (x_1\sigma, x_1\sigma)$ for $\sigma = [U/x_1]$; so if $(x_1, x_1) \in \mathcal{B}$ then $u_2 \models$ NOP.)*

Proposition 41 *A det-first-order grammar $\mathcal{G}$ has a stair-base of width 0 iff $\mathcal{G}$ is s-bounded for some $s \in \mathbb{N}$. If $\mathcal{G}$ is s-bounded then it has a sufficient basis.*

Proof: The first part follows directly from the definitions (if the width is 0 then $\mathcal{G}$ is $g(0)$-bounded). For the second part it suffices to define $\mathcal{B}$ as the set of all sound pairs (E, F) with $\textsc{PresSize}(E, F) \leq s$ (recalling Proposition 38). □

Proposition 42 *If a det-first-order grammar $\mathcal{G}$ has a stair-base (of some width) then it is s-bounded (for some s) and thus has a sufficient basis.*

Proof: Assume a fixed $\mathcal{G}$ which has a stair-base of width $n > 0$, for a fixed function g. By Proposition 41, we are done once we show that $\mathcal{G}$ has also a stair-base of width $n-1$.

So let us consider an arbitrary pair $T_0 \sim U_0$ and some $\alpha \in \textsc{trac}^\omega(T_0)$ which is not $g(0)$-bounded (for T_0, U_0); there are prefixes $w_0, w_1, w_2, \ldots$ of α such that $w_j \models_{(T_0,U_0)} (E_j(T_1, \ldots, T_n), F_j(T_1, \ldots, T_n))$ where $\textsc{PresSize}(E_j, F_j) < g(j)$. (E_0, F_0) is not sound (since α is not $g(0)$-bounded) but $E_0(T_1, \ldots, T_n) \sim F_0(T_1, \ldots, T_n)$ (Proposition 31, Point 1.). There is thus a shortest v exposing an equation for (E_0, F_0) (recall Proposition 28); w.l.o.g. we can assume that the equation is $x_n \doteq H(x_1, \ldots, x_n)$ (where $H \neq x_n$), and thus $w_0 v \models_{(T_0,U_0)} (T_n, H(T_1, \ldots, T_n))$. Since there are only finitely many pairs (E, F) with $\textsc{PresSize}(E, F) < g(0)$, there is some $s \in \mathbb{N}$ determined by $\mathcal{G}$ and $g(0)$ (independent of T_0, U_0, α) such that $|v| < s$ and $\textsc{PresSize}(H) < s$.

By Limit subterm replacement we get $w'_j \models (E'_j(T_1, \ldots, T_{n-1}), F'_j(T_1, \ldots, T_{n-1}))$ for $j = 0, 1, 2, \ldots$, where $w'_j = w_{s+j}$, $E'_j = E_{s+j}[H^{lim_n}/x_n]$ and $F'_j = F_{s+j}[H^{lim_n}/x_n]$. So defining $g'(j) = g(s + j) + f(s)$, for some trivial function f (e.g. $f(s) = 2s$, depending on the definition of the size of graph presentations), shows that $\mathcal{G}$ has a stair-base of width $n-1$ (since our reasoning was independent of T_0, U_0, α). □

Now we aim to show that any det-first-order $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ (in normal form) has a stair-base (of some width). We use further auxiliary notions, recalling $w(X, i)$, M_0, B-Inc, and the d-prefix form ($d = M_0$ in our case).

Definition 43 *Given T and $w \in \mathcal{A}^* \cup \mathcal{A}^\omega$, w* exposes a subterm *of T* in depth d *if there is a prefix u of w such that $P^T_d \xrightarrow{u} x_i$ for some i. (P^T_d is the d-prefix of T.)*
For $T \xrightarrow{w}$, $w \in \mathcal{A}^* \cup \mathcal{A}^\omega$, T sinks by *w if w exposes a subterm of T in depth 1; hence if* T does not sink by *w then $X x_1 \ldots x_m \xrightarrow{w}$ where X is the root nonterminal of T.*

Now we introduce a key ingredient, the notions of left- and right-balancing segments, with the right- and left-balancing pivots; we also use B for ranging over ground terms (which serve as balancing pivots). We assume that the *underlying $\mathcal{G}$ is in normal form* (discussing this issue afterwards).

Definition 44 *A triple (T, B, v) where $|v| = M_0$ is an ℓ-*balancing segment *if $T \sim_{M_0} B$, $T \xrightarrow{v}$ and T does not sink by a proper prefix of v. B is called the* (balancing) pivot *of this segment, an r-*pivot *in this case. The ℓ-*bal-result *ℓ-$\textsc{Res}(T, B, v)$ of this segment is defined as follows: if $T = XT'_1 \dots T'_m$, $Xx_1 \dots x_m \xrightarrow{v} G(x_1, \dots, x_m)$, and $B = P^B_{M_0}(W_1, \dots, W_n)$ then ℓ-$\textsc{Res}(T, B, v)$ is the pair $(G(V_1, \dots, V_m), F(W_1, \dots, W_n))$ where $P^B_{M_0} \xrightarrow{v} F$ and $V_i = F_i(W_1, \dots, W_n)$ where $P^B_{M_0} \xrightarrow{w(X,i)} F_i$ for $i = 1, 2, \dots, m = arity(X)$. ℓ-$\textsc{Res}(T, B, v)$ can be also presented as $(E(W_1, \dots, W_n), F(W_1, \dots, W_n))$ where $E = G(F_1, \dots, F_m)$.*

*An r-*balancing segment *(B, T, v), with the ℓ-*pivot *B and r-$\textsc{Res}(B, T, v)$, is defined symmetrically. We say just "*pivot*" and "*bal-result*" when the side (ℓ or r) follows from context.*

Informally, Proposition 45 captures the following simple idea: if the "left-hand side" (lhs-) term does not sink by a segment of length $M_0{-}1$, so it misses the opportunity to expose a depth-1 subterm by a shortest word, then we can balance, i.e., replace its subterms (in depth 1 originally) by using the rhs-term (r-pivot), achieving a pair with bounded finite heads and the same tails, inherited from the r-pivot. By symmetry the same holds for the case of a non-sinking rhs-term and an ℓ-pivot. In what follows we sometimes leave implicit the parts of the claims which follow by symmetry.

Remark. The normal form assumption on $\mathcal{G}$ is technically convenient (though not really crucial). Definition 44 makes good sense also for $w(X, i) = \varepsilon$ (recall Proposition 18), but Proposition 45 and some later reasoning would be slightly more complicated. An alternative to the normal form assumption would be adding a (harmless) derivation rule in Figure 1 enabling to replace an unexposable subterm arbitrarily.

Proposition 45 *Given an initial pair T_0, U_0, if (T, B, v) is an ℓ-balancing segment and $(V, W) = \ell$-$\textsc{Res}(T, B, v)$ then $u \models (T, B)$ implies $uv \models (V, W)$.*

Proof: Suppose the notation from Definition 44. If $u \models (T, B)$ then by Basic transition rule we get $uv \models (G(T'_1, \dots, T'_m), F(W_1, \dots, W_n))$ and $u(w(X, i)) \models (T'_i, V_i)$ where $V_i = F_i(W_1, \dots, W_n)$, for $i = 1, 2, \dots, m$. Since $|w(X, i)| < M_0 = |v|$, by repeated Subterm replacement we get $uv \models (G(V_1, \dots, V_m), F(W_1, \dots, W_n))$. □

Recalling Observations 20 and 10, we easily observe the following facts.

Observation 46
(1.) The bal-result of an ℓ-balancing segment (T, B, v) is determined by the pivot B, the word v (of length M_0) and by the root nonterminal of T.
(2.) The depth-size of (finite) terms F, G, F_i in the bal-result as in Definition 44 is bounded by $M_0 + \textsc{B-Inc}(M_0)$.

(3.) The number n of tails in the bal-result $(E(W_1, \ldots, W_n), F(W_1, \ldots, W_n))$ has an upper bound determined by $\mathcal{G}$ (since M_0 is determined by $\mathcal{G}$).
(4.) If ℓ-$\textsc{Res}(T, B, v) = (V, W) = (E(W_1, \ldots, W_n), F(W_1, \ldots, W_n))$ where $E = G(F_1, \ldots, F_m)$ as in Definition 44, and $V \xrightarrow{w}$ where w exposes a subterm of V in depth $(1 + \text{B-}\textsc{Inc}(M_0))$ then w necessarily exposes a subterm of V (of the form $F_i(W_1, \ldots, W_n)$) which is reachable from B by some $w(X, i)$ (of length $< M_0$).

We now try to use the possibility of balancing along an infinite $\alpha \in \textsc{trac}^{\omega}(T_0)$, for a pair $T_0 \sim U_0$, to show that α allows a stair-base of width n, with a function g, which are independent of T_0, U_0, α (i.e., with n and g determined just by grammar $\mathcal{G}$). We first observe that if there are only finitely many balancing opportunities then α allows a repeat (recall Definition 40) and the condition is trivial:

Definition 47 *Assume an initial pair $T_0 \sim U_0$ and a fixed $\alpha \in \textsc{trac}^{\omega}(T_0)$. For (the triple u, T, U such that) $u \models_{(T_0,U_0)} (T, U)$, u being a prefix of α, we define the* next ℓ-segment *as the ℓ-balancing segment (T', B, v) for the shortest w (if there is some) such that uwv is a prefix of α and $(T, U) \xrightarrow{w} (T', B)$. The* distance of *this next ℓ-segment is defined as $|w|$. Similarly we define the* next r-balancing segment *for $u \models_{(T_0,U_0)} (T, U)$.*

Proposition 48 *Given T_0, U_0, α as in Definition 47, if there is no next ℓ-segment and no next r-segment for some $u \models_{(T_0,U_0)} (T, U)$, u being a prefix of α, then α has a repeat.*

Proof: Consider performing $\alpha' = u \backslash \alpha = a_1 a_2 a_3 \ldots$ from T; let $T = T_1 \xrightarrow{a_1} T_2 \xrightarrow{a_2} T_3 \xrightarrow{a_3} \cdots$. If there were a (first) segment $T_i \xrightarrow{a_i} T_{i+1} \xrightarrow{a_{i+1}} \ldots \xrightarrow{a_{i+M_0-1}} T_{i+M_0}$ where T_i is a subterm of T but none of $T_{i+1}, T_{i+2}, \ldots, T_{i+M_0}$ is a subterm of T (of T_i, in fact) then we had an ℓ-balancing segment. Hence each T_i is reachable from a subterm of T by a word of length $< M_0$, which means that there are only finitely many different T_i. Similarly for U_i on the rhs. This guarantees a repeat. □

Proposition 49
Any det-first-order grammar $\mathcal{G}$ in normal form has a stair-base (of some width).

Proof: We assume a det-first-order grammar $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ in normal form, a pair $T_0 \sim U_0$ and a fixed $\alpha \in \textsc{trac}^{\omega}(T_0)$; we further write $\models$ instead of $\models_{(T_0,U_0)}$. Assuming that α has no repeat, we show that it has a stair-base of width n, with function g, where n, g are independent of T_0, U_0, α.

We will present α as $u_1 v_1 u_2 v_2 u_3 v_3 \ldots$ where $|v_i| = M_0$, attaching to each v_i a triple $(side_i, T_i, U_i)$ and a pair (T'_i, U'_i) such that $side_i \in \{\ell, r\}$, (T_i, U_i, v_i) is a $side_i$-balancing segment, $(T'_i, U'_i) = side_i$-$\textsc{Res}(T_i, U_i, v_i)$, $(T_0, U_0) \xrightarrow{u_1} (T_1, U_1)$ and $(T'_i, U'_i) \xrightarrow{u_{i+1}} (T_{i+1}, U_{i+1})$ for $i = 1, 2, 3, \ldots$. Hence $u_1 v_1 \ldots u_{i-1} v_{i-1} u_i \models (T_i, U_i)$ and $u_1 v_1 \ldots u_i v_i \models (T'_i, U'_i)$. We note that each v_i has the corresponding pivot B_i, i.e. one of T_i, U_i, depending on $side_i$.

(T_1, U_1, v_1) is defined as the next ℓ-segment or the next r-segment for $\varepsilon \models (T_0, U_0)$; if both exist, the one with the smaller distance is chosen (recall Definition 47), and we prefer ℓ, say, to break ties. This also induces u_1 (where $(T_0, U_0) \xrightarrow{u_1} (T_1, U_1)$).

Suppose $u_1v_1 \dots u_iv_i$ have been defined, and assume that (T_i, U_i, v_i) is an ℓ-segment (so $B_i = U_i$ is an r-pivot; the other case is symmetrical). If for $u_1v_1 \dots u_iv_i \models (T'_i, U'_i)$ there is the next ℓ-segment with the distance at most

$$M_1 = (1 + \textsc{B-Inc}(M_0)) \cdot M_0 \tag{4}$$

then we use this segment to define $u_{i+1}v_{i+1}$ etc.; there was no switch, we have $side_{i+1} = side_i = \ell$. If there is no such "close" ℓ-segment (since the ℓ-side terms keep sinking), we note that a subterm of T'_i in depth $(1+\textsc{B-Inc}(M_0))$ has been exposed by w where $|w| = M_1$ and $u_1v_1 \dots u_iv_iw$ is a prefix of α; let $(T'_i, U'_i) \xrightarrow{w} (T''_i, U''_i)$. Point 4. in Observation 46 implies that T''_i is reachable from B_i, by a word arising from v_iw by replacing a prefix v_iw' with some $w(X, j)$. Here we define $(T_{i+1}, U_{i+1}, v_{i+1})$ as the next ℓ- or r-segment with the least distance for $u_1v_1 \dots u_iv_iw \models (T''_i, U''_i)$ (so w is a prefix of u_{i+1}). This might, but also might not, mean a switch of the pivot side.

Anyway, B_{i+1} is reachable from B_i by w_i where either $w_i = v_iu_{i+1}$ or w_i arises from v_iu_{i+1} by replacing a prefix v_iw' of length $\leq M_0 + M_1$ by a (shorter) word $w(X, j)$. We thus get a *pivot-path*

$$B_1 \xrightarrow{w_1} B_2 \xrightarrow{w_2} B_3 \xrightarrow{w_3} \cdots$$

We note that if some w_i is longer than M_0+M_1, $w_i = w'_iw''_i$ where $|w'_i| = M_0+M_1$, then the "pivot-path sinks" in any segment of w''_i of length M_0, i.e.: for any partition $w''_i = w''_{i1}vw''_{i2}$, $|v| = M_0$, we have $B_i \xrightarrow{w'_iw''_{i1}} W_1 \xrightarrow{v} W_2 \xrightarrow{w''_{i2}} B_{i+1}$ where W_1 sinks by v.

This implies for any $B_1 \xrightarrow{u} V \xrightarrow{\beta'}$ where $u\beta' = \beta = w_1w_2w_3 \dots$ that there is a nonempty prefix u' of β' of length at most

$$M_2 = (M_0 + M_1) + (1 + \textsc{B-Inc}(M_0 + M_1)) \cdot M_0 \tag{5}$$

such that $V \xrightarrow{u'} B_i$ (for some i) or V sinks by u'. (Informally: any segment $V \xrightarrow{w}$ of the pivot path $B_1 \xrightarrow{\beta}$ with length $|w| = M_2$ either contains a pivot B_i or sinks.) Hence if β exposes subterms of B_1 in all depths then infinitely many pivots B_i are equal (since reachable from subterms of B_1 by words of length $\leq M_2$); Point 1. in Observation 46 shows that α would then have a repeat.

So there is the maximal depth d such that $\beta = w_1w_2w_3 \dots$ exposes a (unique) subterm V_1 of B_1 in depth d; hence $B_1 \xrightarrow{u} V_1 \xrightarrow{\beta'}$ where $u\beta' = \beta$ and V_1 does not sink by β'. Let $V_1 = P^{V_1}_{M_0}(T_1, \dots, T_n)$ be the M_0-prefix form of V_1, and let $k \in \mathbb{N}$ be the least such that $B_1 \xrightarrow{u} V_1 \xrightarrow{u'} B_k \xrightarrow{w_k} B_{k+1} \xrightarrow{w_{k+1}} \cdots$ (u' being a prefix of β').

Then pivots B_{k+j}, $j = 0, 1, 2, \dots$, are of the form $G_j(T_1, \dots, T_n)$ where G_j are finite terms in which each occurrence of a variable has depth M_0 at least. Moreover, $\textsc{Depth}(G_j) \leq M_0 + \textsc{B-Inc}(M_2) \cdot (j + 1)$ (by the above "contains a pivot or sinks" fact).

Hence the bal-results (T'_{k+j}, U'_{k+j}) for B_{k+j}, $j = 0, 1, 2, \dots$ are of the form $(E_j(T_1, \dots, T_n), F_j(T_1, \dots, T_n))$ where E_j, F_j are finite terms with the depth-size bounded by $g'(j)$ for some g' determined by the grammar $\mathcal{G}$ (recall Definition 44, Point 2. in Observation 46, and the fact that $M_0, M_1, M_2, \textsc{B-Inc}$ are determined by $\mathcal{G}$). There is thus $g : \mathbb{N} \to \mathbb{N}$ (independent of T_0, U_0, α) such that $\textsc{PresSize}(E_j, F_j) < g(j)$, for $j = 0, 1, 2, \dots$.

Point 3. in Observation 46 thus implies that $\mathcal{G}$ has a stair-base (of some width). □

In fact, we have thus shown the next completeness lemma, and the main theorem.

Lemma 50 *(Completeness)*
For each det-first-order grammar $\mathcal{G}$ in normal form there is a sufficient (sound) basis $\mathcal{B}$.

Theorem 51 *Trace equivalence for deterministic first-order grammars is decidable.*

For deciding $T_0 \overset{?}{\sim} U_0$, an algorithm based on soundness and completeness is clear (using the effective manipulations with graph presentations of regular terms): when we are allowed to generate any finite basis for a given initial pair T_0, U_0 then both questions "$T_0 \not\sim U_0$?", "$T_0 \sim U_0$?" are semidecidable; when verifying $T_0 \sim U_0$, we have to verify all (critical instances of) pairs included in the basis as well.

Remark. By inspecting the proofs we could note that a sufficient basis for a det-first-order grammar (in normal form) is, in fact, computable (since we now know that the value s determined by $\mathcal{G}$ and $g(0)$ in the proof of Proposition 42 is computable) but this computability does not seem much helpful.

Conclusions

The presented proof of the decidability of trace equivalence for det-first-order grammars routinely applies to the dpda language equivalence, as also shown in Appendix 1. The novelty here is the presentation in the framework of first order terms, resulting in a proof which seems technically simpler than the previous ones.

Appendix 2. gives another look at the complexity result by Stirling [6], showing that the framework of first-order terms can be useful there as well.

References

[1] Seymour Ginsburg and Sheila A. Greibach. Deterministic context free languages. *Information and Control*, 9(6):620–648, 1966.

[2] G. Sénizergues. L(A)=L(B)? Decidability results from complete formal systems. *Theoretical Computer Science*, 251(1–2):1–166, 2001.

[3] G. Sénizergues. L(A)=L(B)? a simplified decidability proof. *Theoretical Computer Science*, 281(1–2):555–608, 2002.

[4] Géraud Sénizergues. The equivalence problem for t-turn dpda is co-NP. In *ICALP 2003*, volume 2719 of *LNCS*, pages 478–489. Springer, 2003.

[5] C. Stirling. Decidability of DPDA equivalence. *Theoretical Computer Science*, 255(1–2):1–31, 2001.

[6] C. Stirling. Deciding DPDA equivalence is primitive recursive. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming (ICALP'02)*, volume 2380 of *LNCS*, pages 821–832. Springer-Verlag, 2002 (a full 38-page draft paper available from http://homepages.inf.ed.ac.uk/cps/ [in March 2009]).

Appendix 1.

5 DPDA language equivalence problem presented via trace equivalence for det-first-order grammars

A *deterministic pushdown automaton* (*dpda*) is a tuple $M = (Q, \mathcal{A}, \Gamma, \Delta)$ consisting of finite sets Q of (control) states, $\mathcal{A}$ of actions (or terminals), Γ of stack symbols, and Δ of transition rules. For each pair pA, $p \in Q$, $A \in \Gamma$, and each $a \in \mathcal{A} \cup \{\varepsilon\}$, Δ contains at most one rule of the type $pA \xrightarrow{a} q\alpha$, where $q \in Q$, $\alpha \in \Gamma^*$. Moreover, any pair pA is (exclusively) either *stable*, i.e. having no rule $pA \xrightarrow{\varepsilon} q\alpha$, or *unstable*, in which case there is (one rule $pA \xrightarrow{\varepsilon} q\alpha$ and) no rule $pA \xrightarrow{a} q\alpha$ with $a \in \mathcal{A}$.

A dpda M generates a labelled transition system $(Q \times \Gamma^*, \mathcal{A} \cup \{\varepsilon\}, \{\xrightarrow{a}\}_{a \in \mathcal{A} \cup \{\varepsilon\}})$ where the states are configurations $q\alpha$ ($q \in Q$, $\alpha \in \Gamma^*$). Having our grammars in mind, we view a rule $pA \xrightarrow{a} q\alpha$ as $pAx \xrightarrow{a} q\alpha x$ (for a formal variable x), inducing $pA\beta \xrightarrow{a} q\alpha\beta$ for every $\beta \in \Gamma^*$. The transition relation is extended to words $w \in \mathcal{A}^*$ as usual; we note that $p\alpha \xrightarrow{w} q\beta$ can comprise more than $|w|$ basic steps, due to possible "silent" ε-moves. Each configuration $p\alpha$ has its associated *language* $L(p\alpha) = \{w \in \mathcal{A}^* \mid p\alpha \xrightarrow{w} q\varepsilon \text{ for some } q\}$. The *dpda language equivalence problem* is: given a dpda M and two configurations $p\alpha$, $q\beta$, is $L(p\alpha) = L(q\beta)$?

Remark. It is straightforward to observe that this setting is equivalent to the classical problem of language equivalence between deterministic pushdown automata with accepting states. First, the disjoint union of two dpda's is a dpda. Second, for languages $L_1, L_2 \subseteq \Sigma^*$ we have $L_1 = L_2$ iff $L_1 \cdot \{\$\} = L_2 \cdot \{\$\}$, for an endmarker $\$ \notin \Sigma$; so restricting to prefix-free deterministic context-free languages, accepted by dpda via empty stack, does not mean losing generality.

Each dpda M can be transformed by a standard polynomial-time algorithm so that all ε-transitions are popping, i.e., of the type $pA \xrightarrow{\varepsilon} q$, while $L(pA\alpha)$, for stable pA, keep unchanged. (A principal point is that a rule $pA \xrightarrow{\varepsilon} qB\alpha$ where $qB \xrightarrow{a_1} q_1\beta_1, \ldots,$ $qB \xrightarrow{a_k} q_k\beta_k$ can be replaced with rules $pA \xrightarrow{a_j} q_j\beta_j\alpha$; unstable pairs pA enabling only an infinite sequence of ε-steps are determined and removed.)

It is also harmless to assume that for each stable pA and each $a \in \mathcal{A}$ we have one rule $pA \xrightarrow{a} q\alpha$ (since we can introduce a 'dead' state q_d with rules $q_dA \xrightarrow{a} q_dA$ for all $A \in \Gamma, a \in \mathcal{A}$, and for every 'missing' rule $pA \xrightarrow{a}$ .. we add $pA \xrightarrow{a} q_dA$). $L(p\alpha)$ are unchanged by this transformation. Then $w \in \mathcal{A}^*$ is not enabled in $p\alpha$ iff $w = uv$ where $p\alpha \xrightarrow{u} q\varepsilon$ (for some q), so $u \in L(p\alpha)$, and $v \neq \varepsilon$. This reduces language equivalence to trace equivalence:

$$L(p\alpha) = L(q\beta) \text{ iff } \forall w \in \mathcal{A}^* : p\alpha \xrightarrow{w} \Leftrightarrow q\beta \xrightarrow{w}.$$

Proposition 52 *The dpda language equivalence problem is polynomial-time reducible to the deterministic first-order grammar equivalence problem.*

Proof: Assume an (ε-popping) dpda $M = (Q, \mathcal{A}, \Gamma, \Delta)$ transformed as above (so trace equivalence coincides with language equivalence). We define the first-order grammar $\mathcal{G}_M = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ where $\mathcal{N} = \{pA \mid pA \text{ is } \textit{stable}\} \cup \{\bot\}$; each $X = pA$ gets arity $m = |Q|$, and $\bot$ is a special nullary nonterminal not enabling any action. A dpda configuration $p\alpha$ is transformed to the term $\mathcal{T}(p\alpha)$ defined inductively by rules 1.,2.,3. below, assuming $Q = \{q_1, q_2, \dots, q_m\}$.

1. $\mathcal{T}(q\varepsilon) = \bot$.
2. If $qA \xrightarrow{\varepsilon} q_i$ (qA is unstable) then $\mathcal{T}(qA\beta) = \mathcal{T}(q_i\beta)$.
3. If qA is stable then $\mathcal{T}(qA\beta) = X\,\mathcal{T}(q_1\beta)\dots\mathcal{T}(q_m\beta)$ where $X = qA$.
4. $\mathcal{T}(q_i x) = x_i$.

Rule 4. is introduced to enable the smooth transformation of a dpda rule $pA \xrightarrow{a} q\alpha$, where $a \in \mathcal{A}$, rewritten in the form $pAx \xrightarrow{a} q\alpha x$, to the $\mathcal{G}_M$-rule $\mathcal{T}(pAx) \xrightarrow{a} \mathcal{T}(q\alpha x)$, i.e. to $Xx_1 \dots x_m \xrightarrow{a} \mathcal{T}(q\alpha x)$, where $X = pA$. Thus $\mathcal{R}$ in $\mathcal{G}_M$ is defined (with no ε-moves). We observe easily: if $pA\alpha \xrightarrow{\varepsilon} q\alpha$ (recall that ε-steps are popping) then $\mathcal{T}(pA\alpha) = \mathcal{T}(q\alpha)$; if $pA\alpha \xrightarrow{a} q\beta\alpha$ ($a \in \mathcal{A}$, pA stable) then $\mathcal{T}(pA\alpha) \xrightarrow{a} \mathcal{T}(q\beta\alpha)$. This also implies: if $p\alpha \xrightarrow{w} q\varepsilon$ then $\mathcal{T}(p\alpha) \xrightarrow{w} \bot$. Thus

$$L(p\alpha) = L(q\beta) \text{ iff } (\forall w \in \mathcal{A}^* : p\alpha \xrightarrow{w} \Leftrightarrow q\beta \xrightarrow{w}) \text{ iff } \mathcal{T}(p\alpha) \sim \mathcal{T}(q\beta).$$

We note that $\mathcal{T}(q\alpha)$ can have (at most) $1 + m + m^2 + m^3 + \dots + m^{|\alpha|}$ subterm-occurrences, but the natural finite graph presentation of $\mathcal{T}(q\alpha)$ has at most $1 + m(|\alpha| - 1) + 1$ nodes and can be obviously constructed in polynomial time. □

Appendix 2.

6 A complexity bound

We have, in fact, not fully used the potential of the pivot-path $B_1 \xrightarrow{w_1} B_2 \xrightarrow{w_2} B_3 \xrightarrow{w_3} \cdots$ discussed in the proof of Proposition 49. For showing the existence of a sufficient basis it was sufficient to use just the stair-base subterm V_1 of B_1. We will now explore the idea of the described balancing strategy further, which allows to derive a concrete computable function bounding the length of potential offending words (i.e., the eq-level) when given (a det-first-order grammar $\mathcal{G}$ and) an initial pair (T_0, U_0).

For our aims (in the context of upper complexity bounds), an *elementary function* is a function (of the type $\mathbb{N}^k \to \mathbb{N}$) arising by a finite composition of constant functions, the elementary operations $+, -, \times, \div$, and the exponential operator $\uparrow$, where $a \uparrow n = a^n$.

When we say that a number is *simply bounded*, we mean that there is an elementary function of the size of $\mathcal{G}$ giving an upper bound; e.g., the constants M_0, M_1, M_2, the number of tails in the M_0-prefix form, the depth-size of the heads in any balancing result, etc., are obviously simply bounded.

The "first" nonelementary (hyper)operator is *iterated exponentiation* $\uparrow\uparrow$, also called *tetration*: $a \uparrow\uparrow n = a \uparrow (a \uparrow (a \uparrow (\dots a \uparrow a) \dots))$ where $\uparrow$ is used n-times.

Our analysis will yield the following bound on the length of offending words, which has an obvious algorithmic consequence:

Theorem 53 *For any triple $\mathcal{G}$, T_0, U_0 with the size* InSize *(of a standard presentation), where $\mathcal{G}$ is a det-first-order grammar and (T_0, U_0) is an initial pair such that $T_0 \not\sim U_0$, there is a sequence of actions which is enabled in just one of T_0, U_0 and its length is bounded by $2 \uparrow\uparrow f(\textsc{InSize})$, where f is an elementary function independent of $\mathcal{G}$, T_0, U_0.*

Corollary 54 *Trace equivalence for deterministic first-order grammars can be decided in time (and space) $O(2 \uparrow\uparrow g(\textsc{InSize}))$ for an elementary function g.*

The analogous claims hold for language equivalence of deterministic pushdown automata, as follows from the reduction in the previous section.
We now aim to prove Theorem 53. In the rest of this section we assume a fixed det-first-order grammar $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{R})$ in normal form, if not said otherwise.

Later we will consider a fixed initial pair (T_0, U_0), where $T_0 \not\sim U_0$, and a fixed offending word $\alpha \in \mathcal{A}^*$ for (T_0, U_0). The word α is (now) finite, and our aim is to show an appropriate upper bound on its length which will prove Theorem 53. To this aim, we first show an "upper-bound tool" (Proposition 56 with Corollary 57), and then we demonstrate that the balancing strategy along our finite α (the same strategy as used in the proof of Proposition 49 along the infinite α there) guarantees that the upper-bound tool can be applied to bound the length of α.

We start with noting a possible new type of sound subterm replacement; roughly speaking, if a pair of heads repeats then an equation (if some is exposable) is available for a potential application. (This new subterm replacement serves just for our reasoning, we will not extend the definition of $\models$.) In the next proposition, it might help to imagine that we have $u \models (E(U_1 \dots U_n), F(U_1 \dots U_n))$ and $v \models (E(V_1 \dots V_n), F(V_1 \dots V_n))$ where $|u| < |v|$ and we would like to label v also with $(E(V'_1 \dots V'_n), F(V'_1 \dots V'_n))$, where V'_j arises from V_j by possible replacing of some occurrences of U_n with $H^{lim_n}(U_1, \dots, U_{n-1})$.

Proposition 55 *Assume a pair $(E(x_1, \dots, x_n), F(x_1, \dots, x_n))$ (of regular terms) for which there is a shortest $w \in \mathcal{A}^*$ exposing an equation, w.l.o.g. say $x_n \doteq H(x_1, \dots, x_n)$ ($H \neq x_n$). Let us consider the following three pairs*

$$(E(U_1 \dots U_n), F(U_1 \dots U_n)), \;\; (E(V_1 \dots V_n), F(V_1 \dots V_n)), \;\; (E(V'_1 \dots V'_n), F(V'_1 \dots V'_n)),$$

with eq-levels k_1, k_2, k_3, respectively, where $V_j = G_j(U_n)$ and $V'_j = G_j(H^{lim_n}(U_1, \dots, U_{n-1}))$ (for $j = 1, 2, \dots, n$). Then $k_3 \geq \min\{k_1, k_2\}$, and $k_3 = k_2$ if $k_1 > k_2$.

Proof: We note that $U_n \sim_k H(U_1, \ldots, U_n) \sim_k H^{lim_n}(U_1, \ldots, U_{n-1})$ where $k = max\{k_1 - |w|, 0\}$; hence $V_j \sim_k V'_j$ for $j = 1, 2, \ldots, n$.

Suppose now $k_3 < \min\{k_1, k_2\}$. Then there is an offending prefix w' for the third pair which exposes an equation for E, F (recall Proposition 28); necessarily $|w'| \geq |w|$. We thus have $(E(V'_1, \ldots, V'_n), F(V'_1, \ldots, V'_n)) \xrightarrow{w'} (V'_i, G(V'_1, \ldots, V'_n))$ (or vice versa) for some i and G, where $\textsc{EqLv}(V'_i, G(V'_1, \ldots, V'_n)) = k_3 - |w'| = k'$. But $\textsc{EqLv}(V_i, G(V_1, \ldots, V_n)) \geq k_2 - |w'| \geq k'+1$ and $V_i \sim_{k'+1} V'_i$ (since $k = k_1 - |w| > k'$), which yields a contradiction. Similarly we would contradict the case $k_2 < \min\{k_1, k_3\}$. Hence the claim follows. $\square$

A simple corollary is that if $u \models (E(T), F(T))$ and $v \models (E(e(T)), F(e(T)))$, for a regular term $e = e(x_1)$ (a "1-tail extension"), where $|u| < |v|$, then if v is an offending prefix for (T_0, U_0) then $\textsc{EqLv}(E(e(T)), F(e(T)))$ is independent of T (since $\textsc{EqLv}(E(e(T)), F(e(T))) = \textsc{EqLv}(E(e(H^{lim_1})), F(e(H^{lim_1})))$ for the appropriate H). Hence if we also have $u' \models (E(T'), F(T'))$ and $v' \models (E(e(T')), F(e(T')))$, where $|u'| < |v'| < |v|$ then it is impossible that both v, v' are offending prefixes for (T_0, U_0).

We now show a generalization, which seems related to the Subwords Lemma in [4]. We use a visually more convenient "two-dimensional" notation for (composed) terms: the first rectangle below is a shorthand for $E\sigma_1\sigma_2 \cdots \sigma_r$ where $\sigma_j = [e_1^{i_j}/x_1, \ldots, e_n^{i_j}/x_n]$; it also presupposes that the variables occuring in all terms in the rectangle are from the set $\{x_1, \ldots, x_n\}$.

Given a (head) pair $(E(x_1, \ldots, x_n), F(x_1, \ldots, x_n))$, and n tuples, called *head extensions*, $(e_1^1, \ldots, e_n^1)$, $(e_1^2, \ldots, e_n^2)$, ..., $(e_1^n, \ldots, e_n^n)$, where E, F and $e_j^i = e_j^i(x_1, \ldots, x_n)$ are regular terms, we call (E', F') an *extended head pair* if it can be presented as depicted, for $0 \leq r \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_r \leq n$. We note that there are 2^n such presentations. By (E'_{max}, F'_{max}), called the *maximal pair*, we denote the pair with $r = n$ and $i_1 = 1, i_2 = 2, \ldots, i_n = n$.

$$\boxed{\begin{array}{c} E \\ e_1^{i_1} \ldots e_n^{i_1} \\ e_1^{i_2} \ldots e_n^{i_2} \\ \ldots \\ e_1^{i_r} \ldots e_n^{i_r} \end{array}} \quad \boxed{\begin{array}{c} F \\ e_1^{i_1} \ldots e_n^{i_1} \\ e_1^{i_2} \ldots e_n^{i_2} \\ \ldots \\ e_1^{i_r} \ldots e_n^{i_r} \end{array}}$$

Proposition 56 *Assume* $(E(x_1, \ldots, x_n), F(x_1, \ldots, x_n))$, $(e_1^1, \ldots, e_n^1)$, $(e_1^2, \ldots, e_n^2)$, ..., $(e_1^n, \ldots, e_n^n)$ *as above, and consider a tuple* $T_1, \ldots, T_n$ *(of regular ground terms).*
If $k = \textsc{EqLv}(E'_{max}(T_1, \ldots, T_n), F'_{max}(T_1, \ldots, T_n))$ *is less than* $\textsc{EqLv}(E'(T_1, \ldots, T_n), F'(T_1, \ldots, T_n))$ *for any other extended head pair* (E', F') *then* k *is independent of* $T_1, \ldots, T_n$.

Proof: The claim is trivial for $n = 0$. We prove it for $n > 0$, assuming it holds for $n-1$. If there is no $w \in \mathcal{A}^*$ exposing an equation for E, F then the claim is trivial since $\textsc{EqLv}(E(U_1, \ldots, U_n), F(U_1, \ldots, U_n))$ is independent of $U_1, \ldots, U_n$ (for any $U_1, \ldots, U_n$). So we assume a shortest $w \in \mathcal{A}^*$ exposing an equation for E, F, w.l.o.g. $x_n \doteq H(x_1, \ldots, x_n)$.

Each pair $E'(T_1,\dots,T_n), F'(T_1,\dots,T_n)$ where E', F' is an extended head pair with $i_1 = 1$ gives rise to the depicted pair, by replacing each $e_\ell^{i_j}(x_1,\dots,x_n)$ with $\bar{e}_\ell^{i_j}(x_1,\dots,x_{n-1}) = e_\ell^{i_j}[H^{lim_n}/x_n]$ and by omitting the now superfluous T_n and $e_n^{i_j}$ (for $i_j \neq 1$). This procedure is independent of trees $T_1,\dots,T_n$; these are handled as "black boxes".

$$\boxed{\begin{array}{c} E \\ \bar{e}_1^1 \dots \bar{e}_n^1 \\ \bar{e}_1^{i_2} \dots \bar{e}_{n-1}^{i_2} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_{n-1}^{i_r} \\ T_1 \dots T_{n-1} \end{array}} \boxed{\begin{array}{c} F \\ \bar{e}_1^1 \dots \bar{e}_n^1 \\ \bar{e}_1^{i_2} \dots \bar{e}_{n-1}^{i_2} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_{n-1}^{i_r} \\ T_1 \dots T_{n-1} \end{array}}$$

We thus get 2^{n-1} pairs, with the head pair $(\bar{E}, \bar{F}) = (E(\bar{e}_1^1 \dots \bar{e}_n^1), F(\bar{e}_1^1 \dots \bar{e}_n^1))$ and head extensions $(\bar{e}_1^2,\dots,\bar{e}_{n-1}^2)$, $(\bar{e}_1^3,\dots,\bar{e}_{n-1}^3)$, ..., $(\bar{e}_1^n,\dots,\bar{e}_{n-1}^n)$. Repeated use of Proposition 55 for the triples of the form

$$\boxed{\begin{array}{c} E \\ \bar{e}_1^{i_{\ell+1}} .. \bar{e}_n^{i_{\ell+1}} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_n^{i_r} \\ T_1 \dots T_n \end{array}} \boxed{\begin{array}{c} F \\ \bar{e}_1^{i_{\ell+1}} .. \bar{e}_n^{i_{\ell+1}} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_n^{i_r} \\ T_1 \dots T_n \end{array}} , \boxed{\begin{array}{c} E \\ e_1^{i_1} \dots e_n^{i_1} \\ \dots \\ e_1^{i_{\ell-1}} .. e_n^{i_{\ell-1}} \\ e_1^{i_\ell} \dots e_n^{i_\ell} \\ \bar{e}_1^{i_{\ell+1}} .. \bar{e}_n^{i_{\ell+1}} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_n^{i_r} \\ T_1 \dots T_n \end{array}} \boxed{\begin{array}{c} F \\ e_1^{i_1} \dots e_n^{i_1} \\ \dots \\ e_1^{i_{\ell-1}} .. e_n^{i_{\ell-1}} \\ e_1^{i_\ell} \dots e_n^{i_\ell} \\ \bar{e}_1^{i_{\ell+1}} .. \bar{e}_n^{i_{\ell+1}} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_n^{i_r} \\ T_1 \dots T_n \end{array}} , \boxed{\begin{array}{c} E \\ e_1^{i_1} \dots e_n^{i_1} \\ \dots \\ e_1^{i_{\ell-1}} .. e_n^{i_{\ell-1}} \\ \bar{e}_1^{i_\ell} \dots \bar{e}_n^{i_\ell} \\ \bar{e}_1^{i_{\ell+1}} .. \bar{e}_n^{i_{\ell+1}} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_n^{i_r} \\ T_1 \dots T_n \end{array}} \boxed{\begin{array}{c} F \\ e_1^{i_1} \dots e_n^{i_1} \\ \dots \\ e_1^{i_{\ell-1}} .. e_n^{i_{\ell-1}} \\ \bar{e}_1^{i_\ell} \dots \bar{e}_n^{i_\ell} \\ \bar{e}_1^{i_{\ell+1}} .. \bar{e}_n^{i_{\ell+1}} \\ \dots \\ \bar{e}_1^{i_r} \dots \bar{e}_n^{i_r} \\ T_1 \dots T_n \end{array}}$$

guarantees that the maximal pair (in the new 2^{n-1} pairs) has the same eq-level as the original maximal pair (in the originally assumed 2^n pairs), and this eq-level is less than the eq-level of any other pair. We can thus apply the induction hypothesis. □

For stating an important corollary we introduce the following notion (which assumes a fixed triple $\mathcal{G}, T_0, U_0$). Given a head pair $E(x_1,\dots,x_n), F(x_1,\dots,x_n)$ and head extensions $(e_1^1,\dots,e_n^1)$, ..., $(e_1^n,\dots,e_n^n)$, we say that, for a tuple $T_1,\dots,T_n$, the pair $(E'_{max}(T_1,\dots,T_n), F'_{max}(T_1,\dots,T_n))$ is *saturated on level* $m \in \mathbb{N}$ if for each other extended head pair (E', F') there is u, $|u| < m$, such that $u \models (E'(T_1,\dots,T_n), F'(T_1,\dots,T_n))$.

Corollary 57 *If $u \models (U, U')$ for an offending prefix u for (T_0, U_0) where (U, U') can be presented as a pair $(E'_{max}(T_1,\dots,T_n), F'_{max}(T_1,\dots,T_n))$ saturated on level $|u|$ then there cannot exist an offending prefix v, $|v| \neq |u|$, and a tuple $T'_1,\dots,T'_n$ such that $v \models (V, V')$ where (V, V') can be presented as $(E'_{max}(T'_1,\dots,T'_n), F'_{max}(T'_1,\dots,T'_n))$ saturated on level $|v|$ (where the head pair E, F and the head extensions are the same in both cases).*

We now fix T_0, U_0, where $T_0 \not\sim U_0$, and a (finite) $\alpha \in \mathrm{OW}(T_0, U_0)$, and use the same (balancing) strategy along α as we used along the infinite α in the proof of Proposition 49; i.e., we present α in the appropriate form $u_1 v_1 u_2 v_2 \dots$ as long as possible. We note that we are guaranteed that α does not allow a repeat, since it is offending; if $u \models (T, U)$ for a prefix u of α then $\textsc{EqLv}(T, U) = \textsc{EqLv}(T_0, U_0) - |u|$. Instead of an infinite pivot path we now get a finite pivot path

$$B_1 \xrightarrow{w_1} B_2 \xrightarrow{w_2} \cdots \xrightarrow{w_{k-1}} B_k \,,$$

maybe with $k = 0$ (i.e., with no balancing step at all). But recall the *pivot-path property*: each segment $V \xrightarrow{w}$ of length $|w| = M_2$ either contains a pivot, i.e. $V \xrightarrow{u} B_i$ for a prefix u of w, or sinks, i.e. V sinks by w.

Each pivot B_i, for $i = 1, 2, \ldots, k{-}1$, has the unique subterm V_i (maybe $V_i = B_i$) which is *exposed by a proper prefix of* w_i but none of its proper subterms is thus exposed. This yields the path

$$B_1 \xrightarrow{w_{11}} V_1 \xrightarrow{w_{12}} B_2 \xrightarrow{w_{21}} V_2 \xrightarrow{w_{22}} B_3 \xrightarrow{w_{31}} \cdots \xrightarrow{w_{k-2,2}} B_{k-1} \xrightarrow{w_{k-1,1}} V_{k-1} \xrightarrow{w_{k-1,2}} B_k \tag{6}$$

where w_{i1} can be empty but w_{i2} are nonempty and $|w_{i2}| \leq M_2$. We note that for each segment $V_i \xrightarrow{w_{i2}} B_{i+1} \xrightarrow{w_{i+1,1}} V_{i+1}$ we either have that V_i does not sink by $w_{i2}w_{i+1,1}$ or V_{i+1} is a subterm of V_i.

Definition 58 *We call a subsequence* $(i_0, i_1, \ldots, i_r)$ *of the sequence* $(1, 2, \ldots, k{-}1)$ *a* stair sequence *if for each* $j \in \{0, 1, \ldots, r{-}1\}$ *we have that* V_{i_j} *does not sink by* w *where* $V_{i_j} \xrightarrow{w} V_{i_{j+1}}$ *is the appropriate segment of (6), so* $w = w_{(i_j,2)} w_{i_j+1} \ldots w_{i_{j+1}-1} w_{(i_{j+1},1)}$.
A stair sequence *is* maximal *if it is not a proper subsequence of any other stair sequence.*

Proposition 59 *If* $(i_0, i_1, \ldots, i_r)$ *is a maximal stair sequence then* $V_{i_0}, V_{i_1}, V_{i_2}, \ldots, V_{i_r}$ *can be presented as*

$$V_{i_0} = \boxed{\begin{array}{c} G_0 \\ T_1 \ldots T_n \end{array}}\ , \ V_{i_1} = \boxed{\begin{array}{c} G_1 \\ e^1_1 \ldots e^1_n \\ T_1 \ldots T_n \end{array}}\ , \ V_{i_2} = \boxed{\begin{array}{c} G_2 \\ e^2_1 \ldots e^2_n \\ e^1_1 \ldots e^1_n \\ T_1 \ldots T_n \end{array}}\ , \ \ldots, \ V_{i_r} = \boxed{\begin{array}{c} G_r \\ e^r_1 \ldots e^r_n \\ \ldots \\ e^2_1 \ldots e^2_n \\ e^1_1 \ldots e^1_n \\ T_1 \ldots T_n \end{array}}$$

where G_j *are* M_0*-prefixes (so* n *is simply bounded) and* $\textsc{Depth}(e^i_j) \leq \text{B-}\textsc{Inc}(M_2)$.

Proof: Assuming a maximal stair sequence, we first show that $V_{i_{j+1}}$ is a subterm of V_{i_j+1}: For any $\ell \geq 1$ such that $i_j + \ell < i_{j+1}$ we must have that $V_{i_j+\ell}$ sinks by w' for the appropriate segment $V_{i_j+\ell} \xrightarrow{w'} V_{i_{j+1}}$, which implies that there is ℓ' such that $i_j + \ell < i_j + \ell' \leq i_{j+1}$ where $V_{i_j+\ell'}$ is a subterm of $V_{i_j+\ell}$.

By definition, for the segment $V_{i_j} \xrightarrow{w} V_{i_{j+1}}$ we have $Y x_1 \ldots x_m \xrightarrow{w} F(x_1, \ldots, x_m)$ where Y is the root nonterminal of V_{i_j} and F is not a variable. Recalling the above pivot-path property and the fact that $V_{i_{j+1}}$ is a subterm of V_{i_j+1}, we deduce that $1 \leq \textsc{Depth}(F) \leq 1{+}\text{B-}\textsc{Inc}(M_2)$. This easily implies the claim. (Note that e^i_j can be just a variable. It is also possible that some T_j, e^i_j get obsolete, do not really matter in the respective substitutions.)

□

Proposition 60 *There is an elementary function* g*, independent of* $\mathcal{G}, T_0, U_0, \alpha$*, such that* $|\alpha| \leq g(\textsc{InSize}, \ell)$ *where* ℓ *is the length of the longest stair sequence.*

Proof: For any ground term V which can be presented as $V = F(W_1, \ldots, W_m)$ where F is a finite term and W_i are subterms of T_0 or U_0, let us define $size(V)$ as the least $\textsc{Depth}(F)$ in such presentations; we note that if V' is a subterm of V then $size(V') \leq size(V)$. Since either $T_0 \xrightarrow{u_1} B_1$ or $U_0 \xrightarrow{u_1} B_1$, the above $size$ is well defined for all B_i and V_i; moreover, $size(V_{i+1}) \leq size(V_i) + M_3$ where we put $M_3 = 1 + \text{B-}\textsc{Inc}(M_2)$.

We now show that if $size(V_i) > pM_3$ (for $p \in \mathbb{N}$) then there is a stair sequence $(i_0, i_1, \ldots, i_p)$ such that $i_p = i$: Suppose V_i is a counterexample for the least i and some p; we necessarily have $p \geq 1$. Since V_1 is a subterm of B_1 and B_1 is reachable from a subterm of T_0 or U_0 by less than M_0 moves, we surely have $size(V_1) \leq M_3$; hence $i > 1$. We have $size(V_{i-1}) > (p{-}1)M_3$ and V_i is a subterm of V_{i-1} (since V_{i-1} satisfies the claim for $p-1$ and thus $V_{i-1} \xrightarrow{w} V_i$ necessarily sinks); hence $size(V_{i-1}) > pM_3$. Then $i{-}1 > 1$ and we also have $size(V_{i-2}) > (p{-}1)M_3$ and V_i is a subterm of V_{i-2}, so $size(V_{i-2}) > pM_3$; continuing this reasoning leads to a contradiction.

Hence if ℓ is the length of the longest stair sequence then we have $size(B_i) \leq (\ell+2)M_3$ for each pivot B_i, which gives an elementary bound (in $\textsc{InSize}$ and ℓ) on the length of the pivot path, and thus an elementary bound on $|\alpha|$ as well. □

Thus to prove Theorem 53, it is sufficient to show Proposition 64. We show this by using Corollary 57 and a few combinatorial facts. (The combinatorial facts could be surely found in the literature in some form, e.g. for so called *Zimin words*, but we provide short self-contained versions tailored to our aims.)

Definition 61 *Given an alphabet Σ, the empty word ε is of* type 0*; for $n \geq 1$, a word $w \in \Sigma^*$ is of* type n *if $w = vuv$ for some v of type $n{-}1$ and some u, $|u| \geq 1$. Each word w of type $n > 0$ has a* type-n presentation *given by nonempty words $v_1, v_2, \ldots, v_n$; these give rise to words $w_1, w_2, \ldots, w_n$, where w_i is of type i, as follows: $w_1 = v_1$, $w_2 = w_1 v_2 w_1$ ($= v_1 v_2 v_1$), $w_3 = w_2 v_3 w_2$ ($= v_1 v_2 v_1 v_3 v_1 v_2 v_1$), $\ldots$, $w_n = w_{n-1} v_n w_{n-1}$, where $w_n = w$.*

Proposition 62 *For a type-$(n{+}1)$ presentation of $w \in \Sigma^*$, given by words $v_1, v_2, \ldots, v_{n+1}$, there are words $u_1, u_2, \ldots, u_n$, all beginning with the first symbol of v_1, such that $u_{i_1} u_{i_2} \ldots u_{i_r}$ is a suffix of (the right quotient) w/v_1 for all $1 \leq r \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_r \leq n$.*

Proof: The words $v_1, v_2, \ldots, v_{n+1}$ give rise to $w_1, w_2, \ldots, w_{n+1}$ as in Definition 61. We denote $w'_i = w_i/v_1$, and put $u_i = v_1 v_{i+1} w'_i$ for $i = 1, 2, \ldots, n$ (so $w'_{i+1} = w'_i u_i = w'_i v_1 v_{i+1} w'_i$). Hence

$$\begin{array}{llll}
w'_1 = & \varepsilon & u_1 = & v_1 v_2 \\
w'_2 = & v_1 v_2 & u_2 = & v_1 v_3 v_1 v_2 \\
w'_3 = & v_1 v_2 v_1 v_3 v_1 v_2 & u_3 = & v_1 v_4 v_1 v_2 v_1 v_3 v_1 v_2 \\
\ldots & & \ldots &
\end{array}$$

Since w'_i is a suffix of w'_j for each $j > i$, it is sufficient to show that $u_{i_1} u_{i_2} \ldots u_{i_r}$ is a suffix of w'_{i_r+1}. Since $w'_{i_r+1} = w'_{i_r} u_{i_r}$ and by induction hypothesis we can assume that $u_{i_1} u_{i_2} \ldots u_{i_{r-1}}$ is a suffix of $w'_{i_{r-1}+1}$, and thus of w'_{i_r}, the claim is clear. □

For $h \in \mathbb{N}$, we define function $f_h : \mathbb{N} \to \mathbb{N}$ (recursively) and note the next proposition:

$$f_h(0) = 0, \quad f_h(n+1) = (1 + f_h(n)) \cdot h^{f_h(n)}\,.$$

Proposition 63 *If $|\Sigma| = h$ then each $w \in \Sigma^*$, $|w| \geq f_h(n)$, contains a subword of type n.*

Proof: By induction, using the pigeonhole principle. For $n = 0$ the claim is obvious. Any word w of length $f_h(n+1)$ necessarily contains two occurrences of some (sub)word u of length $f_h(n)$ separated by a nonempty word. Thus $w = v_1 u v_2 u v_3$ where $|v_2| \geq 1$ and $u = u_1 u_2 u_3$ with u_2 of type n; this means that the subword $u_2 u_3 v_2 u_1 u_2$ is of type $n+1$. □

Proposition 64 *There is an elementary function g, independent of $\mathcal{G}, T_0, U_0, \alpha$, such that the length of any stair sequence has an upper bound $2 \uparrow\uparrow g(size(\mathcal{G}))$.*

Proof: Given a maximal stair sequence $(i_0, i_1, \dots, i_r)$ and the presentation of $V_{i_0}, V_{i_1}, V_{i_2}, \dots, V_{i_r}$ as in Proposition 59, let us consider the pivot (sub)sequence $B_{i_0+1}, \dots, B_{i_{r-1}+1}$ (B_{i_j+1} is the first pivot after V_{i_j}). The balancing results with $B_{i_0+1}, \dots, B_{i_{r-1}+1}$ can be presented as

$$\boxed{\begin{array}{c}E_0\\ T_1 \dots T_n\end{array}}\ \boxed{\begin{array}{c}F_0\\ T_1 \dots T_n\end{array}}\ ,\ \boxed{\begin{array}{c}E_1\\ e^1_1 \dots e^1_n\\ T_1 \dots T_n\end{array}}\ \boxed{\begin{array}{c}F_1\\ e^1_1 \dots e^1_n\\ T_1 \dots T_n\end{array}}\ ,\dots,\ \boxed{\begin{array}{c}E_{r-1}\\ e^{r-1}_1 \dots e^{r-1}_n\\ \dots\\ e^2_1 \dots e^2_n\\ e^1_1 \dots e^1_n\\ T_1 \dots T_n\end{array}}\ \boxed{\begin{array}{c}F_{r-1}\\ e^{r-1}_1 \dots e^{r-1}_n\\ \dots\\ e^2_1 \dots e^2_n\\ e^1_1 \dots e^1_n\\ T_1 \dots T_n\end{array}}$$

where $E_i, F_i, e^i_1 \dots e^i_n$ are finite terms whose depth-size is simply bounded (unlike in/around Proposition 56 where we considered general regular terms). Hence the tuples $(E_i, F_i, e^i_1, \dots, e^i_n)$ can be viewed as elements of an alphabet with h elements, where h is bounded by an elementary function of $size(\mathcal{G})$. If $r-1 \geq f_h(n+2)$ then the word

$$(E_{r-1}, F_{r-1}, e^{r-1}_1, \dots, e^{r-1}_n)\ (E_{r-2}, F_{r-2}, e^{r-2}_1, \dots, e^{r-2}_n)\ \dots\ (E_1, F_1, e^1_1, \dots, e^1_n)$$

contains two different occurrences of a subword of type $n+1$, by Proposition 63. By Proposition 62, from this subword we can extract a pair E, F (i.e., the "head-projection" of the first symbol of v_1), and n words ("extension-projections" of $u_1, \dots, u_n$) $(\bar{e}^1_1, \dots, \bar{e}^1_n)$, $(\bar{e}^2_1, \dots, \bar{e}^2_n)$, ..., $(\bar{e}^n_1, \dots, \bar{e}^n_n)$, where each $(\bar{e}^j_1, \dots, \bar{e}^j_n)$ corresponds to the substitution arising by composing the substitutions corresponding to a segment $\boxed{\begin{array}{c}e^{k+\ell}_1 \dots e^{k+\ell}_n\\ \dots\\ e^k_1 \dots e^k_n\end{array}}$ so that: there are $U_1, U_2 \dots, U_n$ determined by the first occurrence of the type $n+1$ subword such that for each tuple $1 \leq i_1 < i_2 < \dots < i_r \leq n$ there is a pair $\boxed{\begin{array}{c}E\\ \bar{e}^{i_1}_1 \dots \bar{e}^{i_1}_n\\ \dots\\ \bar{e}^{i_r}_1 \dots \bar{e}^{i_r}_n\\ U_1 \dots U_n\end{array}}\ \boxed{\begin{array}{c}F\\ \bar{e}^{i_1}_1 \dots \bar{e}^{i_1}_n\\ \dots\\ \bar{e}^{i_r}_1 \dots \bar{e}^{i_r}_n\\ U_1 \dots U_n\end{array}}$ in the above sequence, where the appropriate $(E'_{max}(U_1, \dots, U_n), F'_{max}(U_1, \dots, U_n))$ is saturated on the corresponding level. Similarly for $V_1, V_2, \dots, V_n$ determined by the second occurrence of the type $n+1$ subword. This would yield a contradiction with Corollary 57.

Therefore $r - 1 < f_h(n+2)$. From the definition of f_h we can easily derive $f_h(n+2) \leq h \uparrow\uparrow g_1(n)$ for an elementary function g_1. Since h and n are bounded by elementary functions of $size(\mathcal{G})$, the claim follows. $\square$